

# UNIVERSITA' DEGLI STUDI DI VERONA

*DEPARTMENT OF  
CULTURES AND CIVILISATIONS*

*GRADUATE SCHOOL OF  
HUMANITIES*

*DOCTORAL PROGRAM IN  
PHILOLOGY, LITERATURE AND LINGUISTICS*

*WITH THE FINANCIAL CONTRIBUTION OF  
FONDAZIONE CARIVERONA*

Cycle / year XXIX / 2014

TITLE OF THE DOCTORAL THESIS

A HISTORY OF THE CONCEPT OF PARAMETER IN GENERATIVE GRAMMAR

S.S.D. L-LIN/01

Coordinator: Prof./ssa RAFFAELLA BERTAZZOLI

Tutor: Prof. GIORGIO GRAFFI

Co-tutor: Prof. DENIS DELFITTO

Doctoral Student: Dott. ALESSANDRO RIOLFI

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License, Italy. To read a copy of the licence, visit the web page:

<http://creativecommons.org/licenses/by-nc-nd/3.0/>



**Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



**NonCommercial** — You may not use the material for commercial purposes.



**NoDerivatives** — If you remix, transform, or build upon the material, you may not distribute the modified material.

*A history of the concept of parameter in Generative Grammar*

Alessandro Riolfi

PhD thesis

Verona, 23 May 2017

## ABSTRACT

This thesis traces the history of the concept of parameter in Generative Grammar, from the first steps of the Principles and Parameters model in the late 1970s to the advent of the Minimalist Program (MP), examining how this notion has been implemented both during and after this transition. The analysis carried out in this dissertation starts from the systematization of the so-called “standard theory” of Generative Grammar in *Aspects of the Theory of Syntax* (1965) until the last developments of the MP.

Chapter I offers an overview of the protohistory of the concept of parameter by focusing on the factors, both theoretical and empirical, at the basis of the systematic formulation of this notion in Chomsky (1981a). The theoretical factors are identified with the distinction between *descriptive* and *explanatory adequacy* and Chomsky’s proposed solution to the so-called problem of the *poverty of the stimulus*. The empirical factor consists in the outcome of Rizzi’s and Taraldsen’s pre-parametric inquiries, which shed new light on the systematicity of linguistic variation.

In Chapter II, I examine the individual formulation of the main parameters that were proposed in Generative Grammar within the Government-Binding (GB) Theory of the Eighties. While the parameters at issue are taken from the list that is proposed in Rizzi (2014), in the first part of the chapter they are retrospectively classified according to the specific syntactic property they would refer to in current minimalist theories.

Chapter III focuses on the debate about the concept of parameter which took place during the first decade of the 21<sup>st</sup> century. The first two positions which are discussed are Kayne’s (2000, 2005) microparametric approach, which draws from the idea that parametric variation is located in the lexicon, and Baker’s (2001, 2008a) macroparametric approach, which instead relies on the classical idea that parameters are expressed on principles. These two approaches are then confronted with Newmeyer’s (2004, 2005) criticism, which points out their descriptive and theoretical flaws. This chapter ends with the presentation of the parametric model proposed by Roberts & Holmberg (2010), which overcomes the limitations of micro- and macro-parameters by combining a lexically-based, microparametric view of linguistic variation with the idea that parametric variation is an emergent property of the interaction of UG, primary linguistic data, and third-factor considerations.

Chapters IV and V evaluate the classical parameters of the GB Theory which still play a role in current generative theory. Chapter IV reviews the null subject parameter, the

V-to-T movement parameter, the polysynthesis parameter, and the overt vs covert wh-movement parameter, while Chapter V is devoted to the history of the head-complement parameter. While on the one hand null subject, V-to-T, and polysynthesis can be reconciled with Roberts & Holmberg's theory, which is based on the assumption that the locus of parameters is the functional lexicon, on the other it is argued that wh-movement and head-directionality pertain to the A-P interface, as envisioned by Berwick & Chomsky (2011). The picture emerging from this analysis highlights that the nature of parametric variation is twofold: syntactic and post-syntactic. This has an interesting consequence on the duality between head-movement and phrasal movement, as only in narrow syntax heads are observed to move, with XPs being linearized post-syntactically.

## **ACKNOWLEDGEMENTS**

The writing of this PhD thesis has not been an easy task. Surprisingly, this is also true of this acknowledgments section, which would easily rival the length of this whole work if I could thank all the people who have been supporting and helping me during the last three years.

First of all, I would like to thank my supervisors, prof. Giorgio Graffi and prof. Denis Delfitto. In particular, prof. Graffi has been following my work ever since my MA thesis, and I am very grateful to him not only for always being supportive and for helping me with my writings, but also and especially for transmitting me his passion for linguistics and its history. During our meetings, his knowledge of the subject and his sharp observations have stimulated my curiosity and have pushed me further and further in developing my research. I am also grateful to the Department of Cultures and Civilisations of the University of Verona, whose activities and seminars have given me food for thought and have stimulated my growth as a scholar.

Secondly, I am very grateful to prof. Luigi Rizzi and prof. Uri Shlonsky. During my experience as a visiting student at the University of Geneva, their insights and suggestions on my research project have significantly contributed to the final shape of this thesis.

Finally, but not least, I would like to thank my thesis evaluators, prof. Frederick Newmeyer and prof. Uri Shlonsky, for their useful comments and their feedback.



## TABLE OF CONTENTS

<b>Chapter 1 – The birth of the concept of “parameter” in Generative Grammar and its development until <i>Lectures on Government and Binding</i> (1981a)</b>	<b>11</b>
1.1 – The theoretical foundations of Generative Grammar	11
1.1.1 – The research program in <i>Aspects of the Theory of Syntax</i> (1965)	12
1.1.2 – The paradox of language learning: the logical problem of the “poverty of the stimulus”	16
1.2 – Prehistory of the term “parameter”	19
1.2.1 – The Extended Standard Theory and the search for linguistic universals	19
1.2.2 – The term “parameter” in <i>Conditions on Rules of Grammar</i> (1976)	22
1.3 – Generative Grammar’s development in the second half of the Seventies and the first explicit parameters	25
1.3.1 – The Subjacency Condition and the Nominative Island Condition	25
1.3.2 – From a language-particular approach to a comparative approach in the investigation of UG	29
1.3.3 – The first example of parametrization: Rizzi and the principle of Subjacency	30
1.3.4 – Taraldsen and the Null Subject parameter	32
1.3.5 – Toward the systematic formulation of the P&P model	35
1.4 – The concept of parameter as an epistemological necessity	36
1.4.1 – The concept of parameter in the “Pisa Lectures”	36
1.4.2 – Language typology in the P&P model: core grammar and periphery	39
1.4.3 – The concept of parameter and language acquisition	41
1.5 – Conclusions	43
 <b>Chapter 2 – The formulation of the main parameters of the Government-Binding Theory</b>	 <b>45</b>
2.1 – A retrospective classification of the parameters of the Government-Binding Theory	45
2.1.1 – Locality, Case-assignment, Merge and Linearization Parameters	46
2.1.2 – Some notes on Spellout Parameters	47
2.2 – The formulation of the first explicit parameters	49

2.2.1 – The parameterization of Subjacency extended: the bounding nodes in French .....	50
2.2.2 – The formulation of the Null Subject Parameter .....	57
2.2.3 – The lexical parameter of S'-deletion .....	71
2.3 – Some theoretical remarks on the onset of the parametric enterprise .....	74
2.4 – Some notable systematic differences in the treatment of cross-linguistic variation .....	80
2.4.1 – Cross-linguistic variation at the syntax-semantics interface: a difference in the locus of wh-movement.....	80
2.4.2 – Preposition stranding and V-P reanalysis .....	94
2.4.3 – The complementary distribution of finite verb and lexical complementizer in verb-second languages .....	104
2.5 – The main parameters of the Government-Binding Theory from the mid-Eighties onwards.....	108
2.5.1 – The parameterization of the Projection Principle .....	108
2.5.2 – The parameter of abstract incorporation .....	122
2.5.3 – The parametrization of the notion of governing category .....	136
2.5.4 – The verb movement parameter.....	146
2.5.5 – The parameter of nominative case assignment .....	157
2.6 – Conclusion .....	164
<b>Chapter 3 – The development of the concept of parameter in Minimalism .....</b>	<b>167</b>
3.1 – Introduction .....	167
3.2 – The parametric approach to language variation in the Government-Binding Theory .....	168
3.2.1 – From parametrized principles to lexical parameters .....	168
3.2.2 – Parameters and linguistic typology in the Government-Binding Theory.....	169
3.3 – The parametric approach to language variation in Minimalism .....	171
3.3.1 – Kayne's microparametric approach to language variation .....	171
3.3.1.1 – Kayne's argument for parametric theory .....	174
3.3.1.2 – Microparametric syntax as an alternative to macroparameters.....	176
3.3.2 – Baker's Parameter Hierarchy.....	178
3.3.2.1 – Baker's argument for parametric theory.....	181
3.3.2.2 – The distinction between micro- and macroparameters .....	183



3.3.2.3 – A revaluation of the macroparametric approach .....	185
3.4 – Against a parametric approach to typological variation .....	190
3.4.1 – Against a generative account for typological generalizations .....	190
3.4.2 – Newmeyer’s critique of parameters .....	191
3.5 – Some perspectives on the debate about parametric theory .....	197
3.5.1 – The crucial difference between the conceptions of parameter in the microparametric and the macroparametric approach and their relation to explanatory adequacy .....	197
3.5.2 – Macro- vs micro-parameters: a terminological and epistemological issue ....	200
3.6 – The state of the art of the parameter-setting model for typological variation .....	204
3.6.1 – Against a rule-based model of cross-linguistic variation .....	205
3.6.2 – Roberts and Holmberg’s hierarchical model of parametric variation.....	209
3.7 – Conclusion .....	215

## **Chapter 4 – The main parameters of the GB Theory in current generative theory ..217**

4.1 – Some preliminary considerations on Rizzi’s (2014) list with respect to modern generative theory .....	217
4.1.1 – The parametrization of bounding nodes, the Barriers framework and the Phase Impenetrability Condition .....	218
4.1.2 – A minimalist account of long-distance anaphora: the logophoric use of reflexives .....	220
4.1.3 – The abandonment of the notions of typological non-configurationality and government in Generative Grammar .....	222
4.2 – The main GB parameters in Minimalism .....	224
4.2.1 – The null subject parameter .....	225
4.2.2 – The V-to-T movement parameter.....	233
4.2.3 – The V-to-C movement parameter .....	242
4.2.4 – The polysynthesis parameter.....	247
4.2.5 – The overt vs covert wh-movement parameter.....	255
4.2 – Conclusion .....	263

## **Chapter 5 – The head-complement parameter .....265**

5.1 – The first explicit parameters of word order .....	265
5.1.1 – The parametrization of recursion directionality .....	265

5.1.2 – The elimination of the categorial component of the base and the postulation of the head-complement parameter.....	270
5.1.3 – A parameterization of head-directionality, theta-role assignment, Case-assignment, and predication.....	276
5.2 – The crisis of the head-complement parameter .....	285
5.2.1 – The antisymmetry of syntax.....	285
5.3 – Some recent proposals regarding head-directionality .....	293
5.3.1 – A minimalist reformulation of the head-complement parameter.....	296
5.3.2 – Head directionality, narrow syntax, and the Final-over-Final Constraint.....	302
5.3.3 – An experimental perspective on linearization.....	311
5.4 – Conclusion .....	320
<b>Chapter 6 – Some concluding remarks on parametrization.....</b>	<b>323</b>
<b>REFERENCES .....</b>	<b>327</b>

## Chapter I

### **The birth of the concept of “parameter” in Generative Grammar and its development until *Lectures on Government and Binding* (1981a)**

#### 1.1 – The theoretical foundations of Generative Grammar

The theoretical framework founded by Noam Chomsky and known as Generative Grammar has been one of the most productive linguistic theories since its birth in the late Fifties. While pursuing its ultimate goal of investigating the very nature of human language, the necessity of answering such questions as «what constitutes knowledge of a language, how does such knowledge develop and how is such knowledge put to use» has urged generative linguists, and especially Chomsky, to delve not only into the technical description of more and more grammar systems to expand their collection of linguistic data, but also into the theoretical foundations of linguistics itself (Chomsky 1981b, p. 32). This constant effort has characterised Chomskian linguistics since its very beginning, and its effects have played a big part in allowing Generative Grammar to develop in the course of time as a scientific research program.

Although the cornerstones of Generative Grammar have never really changed throughout the years, Chomsky himself has never been shy of re-discussing, and in some cases even questioning, his own theoretical assumptions. This behaviour, which could erroneously be interpreted as a sign of weakness as well as inconsistency, actually derives from the inherent need of developing and strengthening the status of modern linguistics as an empirical science. The approach of Generative Grammar to language description and analysis is in fact strictly deductive: starting from a general hypothesis, which in this case is precisely the innateness of language faculty, Chomsky's inquiry proceeds by formulating specific speculations in a form that can conceivably be falsified by a test on observable data, derived in turn from linguistic analysis. As long as these speculations are confirmed by empirical evidence, the theoretical model works and it is assumed as valid as it is corroborated by its own predictions. On the other hand, once the data run contrary to these predictions or, from a conceptual perspective, the theory itself exhibits unnecessary redundancies, the hypothesis is amended or abandoned. An historical analysis of the development of Generative Grammar is thus extremely important to fully understand its theoretical steps and to correctly evaluate the progressive efforts of

Chomsky and his associates through the different phases which characterised this theoretical framework from its very beginning to this day.

### 1.1.1 – *The research program in Aspects of the Theory of Syntax (1965)*

Despite the profound evolution of Generative Grammar's conceptual frameworks, the theoretical foundations of Chomsky's syntactic theory have practically remained the same as those outlined in *Aspects of the Theory of Syntax* (Chomsky 1965), the seminal work in which the original model of Generative Grammar (the phase which has been known as *the standard theory*) had been systematically laid out. Although this book was published eight years after Chomsky's first published book *Syntactic Structures* (1957), it represents a reference point for any generative linguists since it «summarized the work of the decade between the mid 1950s and the mid 1960s and reshaped it in a very systematic model» (Graffi 2001, p. 350).

The first chapter of the *Aspects*, aptly named *Methodological Preliminaries*, is particularly important since it sets forth the fundamentals of Chomsky's linguistic research program. The first aspect to be clarified is what the term “generative grammar” as used by Chomsky refers to. According to Chomsky, a generative grammar is a theory of language which is not merely concerned about the taxonomic description of neither a specific language nor a set of languages, but whose primary aim is the explicit «description of the ideal speaker-hearer's intrinsic competence» (Chomsky 1965, p. 4), which in turn is meant as «a system of rules that can iterate to generate an indefinitely large number of structures» (Chomsky 1965, pp. 15-16):

A grammar of a language purports to be a description of the ideal speaker-hearer's intrinsic competence. If the grammar is, furthermore, perfectly explicit – in other words, if it does not rely on the intelligence of the understanding reader but rather provides an explicit analysis of his contribution – we may (somewhat redundantly) call it a *generative grammar*. (Chomsky 1965, p. 4)

This aspect had already been highlighted in one of the earliest Chomsky's works, namely *The Logical Structure of Linguistic Theory* (LSLT) which, although it was published in 1975, had been written in the mid Fifties. In LSLT great emphasis is placed on the capacity of a grammar to *generate* a potentially infinite set of well-formed sentences by

means of «an intuitive sense of grammaticalness» inherently possessed by each native speaker (Chomsky 1975a, p. 95). Being the notion of “grammaticalness” an intuitive reality rather than extrinsic, «the set of grammatical sentences cannot» simply «be identified with the linguist’s corpus of observed sentences» (Chomsky 1975a, p. 129), but instead coincides with «the speaker’s ability to project his past linguistic experience to form new sentences», that is, to generate a potentially infinite set of well-formed expressions (Chomsky 1975a, p. 132):

The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance. Hence, in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior. (Chomsky 1965, p. 4)

In these terms, within the *Aspects* linguistic competence is assumed to be a strictly psychological concept. This psychological interpretation of linguistic theory is reaffirmed not only by making «a fundamental distinction between *competence* (the speaker-hearer’s knowledge of his language) and *performance* (the actual use of language in concrete situations)» (*ibidem*), but also by regarding the account of the inherent connection between language and mind as the primary means of evaluating the adequacy of two or more candidate grammars. As Chomsky notes, here the term “grammar” is used with a «systematic ambiguity»: in order to refer, on the one hand, to the mentally represented system of knowledge attained by the ideal speaker-hearer and which represents his linguistic competence, and on the other hand, to the theory proposed by the linguist in order to account for this psychological system (Chomsky 1965, p. 25). The study of grammar, understood in this way, imposes to the linguist to choose, among the multiple possible “theories of language”, the one which adheres the most to the mental reality of grammar:

To facilitate the clear formulation of deeper questions, it is useful to consider the abstract problem of constructing an “acquisition model” for language, that is, a theory of language learning or grammar construction. Clearly, a child who has learned a language has developed an internal representation of a system of rules that determine how sentences are to be formed, used, and understood. Using the term “grammar” with a systematic ambiguity (to refer, first, to the native speaker’s internally represented “theory of his language” and, second, to the linguist’s account of this), we can say that the child has developed and

internally represented a generative grammar, in the sense described. He has done this on the basis of observation of what we may call *primary linguistic data*. (Chomsky 1965, pp. 24-25)

Based on such conditions, the central aim of linguistic theory is precisely to account for the process of linguistic acquisition. In fact, with regard to the first of the two aforementioned definitions of grammar, the child himself has to be able to construct his own "theory of language" from among a set of multiple possible alternatives. According to Chomsky, this clearly suggests that, «as a precondition for language learning, he must possess, first, a linguistic theory that specifies the form of the grammar of a possible human language, and, second, a strategy for selecting a grammar of the appropriate form that is compatible with the primary linguistic data» (Chomsky 1965, p. 25):

It seems clear that many children acquire first or second languages quite successfully even though no special care is taken to teach them and no special attention is given to their progress. It also seems apparent that much of the actual speech observed consists of fragments and deviant expressions of a variety of sorts. Thus it seems that a child must have the ability to "invent" a generative grammar that defines well-formedness and assigns interpretations to sentences even though the primary linguistic data that he uses as a basis for this act of theory construction may, from the point of view of the theory he constructs, be deficient in various respects. (Chomsky 1965, pp. 200-201, n. 14)

With regard to the «respects in which one can speak of "justifying a generative grammar"» (Chomsky 1965, p. 26), Chomsky postulates two different levels according to which a linguistic theory can be evaluated. From a purely descriptive perspective, the linguist's task is to give a correct account of the intrinsic competence of the idealised native speaker. This corresponds to the level of *descriptive adequacy*: according to this notion, «a linguistic theory is descriptively adequate if it makes a descriptively adequate grammar available for each natural language» (Chomsky 1965, p. 24). On such terms, the linguist can meet this condition by formulating a system of rules whereby the linguistic facts occurring in a given language are systematically predicted. The grammar is therefore justified on purely empirical grounds or, as Chomsky says, on *external* grounds (cf. Chomsky 1965, p. 27). However, «although even descriptive adequacy on a large scale is by no means easy to approach», according to Chomsky «it is crucial for the productive development of linguistic theory that much higher goals than this be pursued» (Chomsky 1965, p. 24). This higher goal is represented by *explanatory adequacy*, which requires a

linguistic theory to succeed «in selecting a descriptively adequate grammar on the basis of primary linguistic data», that is, a theory which effectively explains how the child develops a system of knowledge of his native language starting from the examples of linguistic performance he is exposed to (Chomsky 1965, p. 25). Considering the psychological reality of linguistic theory, every hypothesis on the nature of linguistic competence corresponds to an hypothesis on the nature of human mind. Therefore, the primary task of generative grammar becomes that of reconciling language description with those specific and innate mechanisms which are directly responsible for language acquisition:

On a much deeper and hence much more rarely attainable level (that of explanatory adequacy), a grammar is justified to the extent that it is a *principled* descriptively adequate system, in that the linguistic theory with which it is associated selects this grammar over others, given primary linguistic data with which all are compatible. In this sense, the grammar is justified on *internal* grounds, on grounds of its relation to a linguistic theory that constitutes an explanatory hypothesis about the form of language as such. The problem of internal justification - of explanatory adequacy - is essentially the problem of constructing a theory of language acquisition, an account of the specific innate abilities that make this achievement possible. (Chomsky 1965, p. 27)

According to Chomsky, the criterion of shaping a linguistic theory which can not only correctly predict a set of linguistic phenomena, but whose framework closely adheres to those general principles underlying the nature of language meant as an innate mental faculty, can really allow linguistics to develop as a scientific theory. This higher benchmark is actually determinant in strengthening linguistic theory as it allows the linguist to select, among two or more conflicting grammars on a par with each other as far as descriptive adequacy is concerned, the one which is more justified on *internal* grounds than the other, that is, on grounds of its relation to those principles which provide an answer to how the child develops his own linguistic competence. Although a purely descriptive grammar may still seem alluring, it does not provide any explanation «concerning the universal properties that determine the form of language» (Chomsky 1965, p. 35). In fact, its predictive power is based on a mere generalisation rather than on a principled theory. Therefore, it «provides no answer to the [...] question: How does the child come to know that the facts are as specified in the descriptively adequate grammar?» (Chomsky 1981b, p. 37):

Clearly, it would be utopian to expect to achieve explanatory adequacy on a large scale in the present state of linguistics. Nevertheless, considerations of explanatory adequacy are

often critical for advancing linguistic theory. Gross coverage of a large mass of data can often be attained by conflicting theories; for precisely this reason it is not, in itself, an achievement of any particular theoretical interest or importance. As in any other field, the important problem in linguistics is to discover a complex of data that differentiates between conflicting conceptions of linguistic structure in that one of these conflicting theories can describe these data only by *ad hoc* means whereas the other can explain it on the basis of some empirical assumption about the form of language. (Chomsky 1965, p. 26)

To summarise briefly, in the very first part of the *Aspects* there are two fundamental concepts which still characterise Chomsky's research program to this day. First, there is the idea of language as an innate mental faculty, according to which generative linguistics is primarily concerned with shedding light on the nature of linguistic competence, the internally-represented grammar shared by all the native speakers of a given language. Second, being generative grammar the study of an actual mental faculty, linguistic theory imposes the linguist to construct a grammar not only of descriptive value, but that is able to account for that specific innate endowment which allows language acquisition.

### *1.1.2 – The paradox of language learning: the logical problem of the “poverty of the stimulus”*

As pointed out above, Generative Grammar is a linguistic theory whose primary aim is to develop a formal apparatus which can account for every well-formed linguistic expression in a given language and, at the same time, can produce an infinite set of sentences by means of a limited set of rules and functional elements. Hence, here the word “generative” has a two possible meanings. If it is related to the creative processes of language, “generative” means *productive* since this linguistic theory provides «a real understanding of how a language can (in Humboldt's words) "make infinite use of finite means"» (Chomsky 1965, p. 8). With regard to the description of the speaker's linguistic knowledge, this term has the meaning of *explicit*, since it aims at making his “intuitive sense of grammaticality”, that is the implicit properties of his internalised grammar, explicit (cf. Graffi 2008, p. 10). However, because of its psychological implications, linguistic theory cannot exempt itself from taking into account the apparent paradox represented by the relation between the process of language acquisition and the so-called problem of the *poverty of the linguistic stimulus*. According to Chomsky, in fact, the most striking aspect of language acquisition is that the primary linguistic data to which the child



has access, even in the best-case scenario, cannot be sufficient to explain the level of proficiency he is bound to reach in his language once he has become a mature native speaker:

The child who acquires a language in this way of course knows a great deal more than he has "learned." His knowledge of the language, as this is determined by his internalized grammar, goes far beyond the presented primary linguistic data and is in no sense an "inductive generalization" from these data. (Chomsky 1965, pp. 32-33)

This logical problem led Chomsky to postulate the existence of an innate «language-acquisition device», often abbreviated to "LAD", «capable of utilizing such primary linguistic data as the empirical basis for language learning» (Chomsky 1965, p. 32) and which, according to Generative Grammar's psychological interpretation, «is only one component of the total system of intellectual structures that can be applied to problem solving and concept formation» – in this case, the task of constructing of a grammar (Chomsky 1965, p. 56). In order to really account for the development of a native speaker's linguistic competence, as Chomsky writes:

This device must search through the set of possible hypotheses [...] and must select grammars that are compatible with the primary linguistic data [...]. The device would then select one of these potential grammars [...]. The selected grammar now provides the device with a method for interpreting an arbitrary sentence [...]. That is to say, the device has now constructed a theory of the language of which the primary linguistic data are a sample. The theory that the device has now selected and internally represented specifies its tacit competence, its knowledge of the language. (Chomsky 1965, p. 32)

This human-specific cognitive structure, later referred to as "universal grammar" (UG), has been described by Chomsky as a sort of "black box" which takes primary linguistic data as its input and produces a language-specific grammar as its output (cf. Chomsky 1981b, pp. 34-35). Since this input-output system, being a mental reality, is not directly observable, for the sake of explanatory adequacy the task of the generative linguist is that of determining the nature of this device considered to underlie language acquisition by formulating hypotheses on the basis of the primary linguistic data associated with each grammar:

Much information can be obtained about both the primary data that constitute the input and the grammar that is the "output" of such a device, and the theorist has the problem of determining the intrinsic properties of a device capable of mediating this input-output relation. (Chomsky 1965, p. 47)

This theoretical advancement, however, would truly represent «the construction of a reasonable acquisition model» only on condition that linguistic theory managed «to reduce the class of attainable grammars compatible with given primary linguistic data», namely, the set of grammars that the child is naturally endowed and which represent his initial linguistic hypotheses (Chomsky 1965, p. 35). On the one hand, the speed at which a child acquires such a complex construct of rules as his native tongue, especially taking into account the scattered and relatively scarce linguistic input he is exposed to, clearly implies the existence of a limited set of core properties, common to all languages, which restrict the class of possible grammars and without which such a task would be theoretically impossible:

A theory of linguistic structure that aims for explanatory adequacy incorporates an account of linguistic universals, and it attributes tacit knowledge of these universals to the child. It proposes, then, that the child approaches the data with the presumption that they are drawn from a language of a certain antecedently well-defined type, his problem being to determine which of the (humanly) possible languages is that of the community in which he is placed. Language learning would be impossible unless this were the case. (Chomsky 1965, p. 27)

On the other hand, according to Chomsky, «the existence of deep-seated formal universals [...] implies that all languages are cut to the same pattern» (Chomsky 1965, p. 30). The fact that all languages share a common core of basic properties – which can thus be regarded as *universal* – suggests the existence of an actual limit to linguistic variation: a limit deriving from the nature of the human brain itself, and whose role is to provide an innate and universal template which represents the basis not only of every *existing* language, but of every *possible* language:

Consequently, the main task of linguistic theory must be to develop an account of linguistic universals that, on the one hand, will not be falsified by the actual diversity of languages and, on the other, will be sufficiently rich and explicit to account for the rapidity and uniformity of language learning, and the remarkable complexity and range of the generative grammars that are the product of language learning. (Chomsky 1965, pp. 27-28)

For this purpose, the next step of Chomsky's research consisted in elaborating a series of constraints on both the form and the applicability of grammatical rules. While the idea that «the critical factor in the development of a fully adequate theory is the limitation of the class of possible grammars» had been already stated in the *Aspects* (Chomsky 1965, p. 61), it wasn't until the early Seventies that this part of Generative Grammar's research program was put into practice and, then, eventually led to the development of the Principles and Parameters (P&P) model.

## 1.2 – Prehistory of the term “parameter”

Since the systematization of the standard model, Generative Grammar has undergone many changes in the types of rules and representations used to reconcile the formal description of individual languages and the more general quest for linguistic universals. In the period from the middle Sixties until now, the syntactic theory founded by Chomsky has been known by different names, each one reflecting a distinct theoretical stage of its continuous development: in the Seventies, “Extended Standard Theory” (EST); in the Eighties, “Government-Binding Theory” (GB-Theory), or “Principles and Parameters Theory” (P&P); finally, from the early Nineties until now, “Minimalist Program” (MP). Although these labels represent, in chronological order, «the three different phases (until now) of the Chomskian program» (Graffi 2001, p. 425), EST, P&P and MP share the same programmatic purpose outlined in *Aspects of the Theory of Syntax*, that is, to «account for the rapidity and uniformity of language learning, and the remarkable complexity and range of the generative grammars that are the product of language learning» (Chomsky 1965, p. 28). In this chapter I will show how the first use of the term “parameter”, although still quite far from assuming the more complex and specific meaning it will be given in the P&P model, related to the specific theoretical context represented by the EST phase.

### 1.2.1 – *The Extended Standard Theory and the search for linguistic universals*

In the decade which coincides with the phase of Generative Grammar known as EST, Chomsky's personal research specifically focused on the identification of those innate universals which, according to the theoretical assumptions laid out in *Aspects of the Theory of Syntax*, would make language acquisition possible. This period was

characterized by two main innovations over the standard model. First, from this point forward, in Chomsky's works there has been an added emphasis on the biological character of Universal Grammar, which is not meant to be a purely mental reality but an actual genetically determined apparatus on a par, in principle, with the other human body organs:

The class of possible human languages is, I assume, specified by a genetically determined property, apparently species-specific in important respects. Any proposed linguistic theory – in particular, EST – may be regarded as an attempt to capture this property, at least in part. Thus a linguistic theory may be understood as a theory of the biological endowment that underlies the acquisition and use of language; in other terms, as a theory of universal grammar (UG), where we take the goal of UG to be the expression of those properties of human language that are biologically necessary. (Chomsky 1977b, p. 2)

Second, Generative Grammar has definitely relied on the idea that linguistic universals, now meant as the components of the highly specialized biological endowment known as UG, do not consist in specific categories or rules, but should be regarded as *conditions* on the format of the grammar of *any* human language (cf. Graffi 2010, p. 416). From this perspective, Chomsky's position on the necessary limitation of possible grammars is clear: the main task of linguistic theory is to limit the generative power of grammars by means of defining the restrictions on the functioning and on the format of syntactic rules. The advantage of this approach over a merely rule-based one is that a limited set of very abstract principles is more appealing, in terms of explanatory adequacy, than a broad set of detailed descriptive rules. On the one hand, a small number of general conditions on rules allows linguistic theory to do without rules which are either too complex or too specific to be ascribed to an innate system governing language acquisition. According to Chomsky, in fact, even without assuming that «the peculiar restrictions on the applicability of rules would have to be built into the rules themselves», the formulation of «general principles that would constrain the application of rules» would «permit the rules themselves to be of quite a simple sort, since many of their detailed properties are, in effect, “factored out”» (Chomsky 1977b, pp. 19-20):

Note again that a crucial contribution of a theory of conditions on rules is that if successful, it makes it unnecessary for individual rules to be richly articulated. While conditions on rule application do not in themselves restrict the class of accessible grammars, they contribute significantly to this end in an indirect manner, by permitting the class of possible rules to be

sharply restricted. In this way a theory of conditions on rules can contribute to a solution of the fundamental problem of accounting for the growth of language, what is called (with misleading connotations, I believe) "language learning" (Chomsky 1977b, p. 20)

On the other hand, the high degree of generality and abstractness which is the prerequisite to such universal conditions is assumed to be the proof of their innate nature. First, their very general scope ensures their potential applicability to every natural language and, conversely, their universality. Second, their elusive nature firmly excludes the possibility that they could be acquired from the outside, since it makes them impossible to acquire through an explicit parental teaching or a set of inductive hypotheses put forward on the basis of primary linguistic data. According to Chomsky, only when linguistic theory formulates such principles that, while having a clear predictive power, can be plausibly assumed to not be learnable because of their abstractness, «we gain some understanding of the innate factors that endow the human language faculty with its remarkable properties» (Chomsky 1977b, pp. 15-16). Moreover, the more these theoretical assumptions are far from the specific grammatical phenomena they indirectly govern, the broader descriptive range they will cover in their restrictive scope. Therefore, the postulation of such general principles would allow linguistic theory not only to fulfil its ultimate goal of explanatory adequacy, but also to account for a greater number of linguistic differences than by using merely descriptive composite rules, thus approaching descriptive adequacy on a much larger scale:

In my personal opinion, it is questions and issues such as these that make the study of language intellectually interesting. That is, at a sufficient level of depth and abstractness of theory, we can expect to discover that small modifications in theoretical assumptions will have varied and complex effects on predicted phenomena. If the predictions are verified, and furthermore, if it cannot plausibly be maintained that the principles have been learned, we can reasonably conclude that we are obtaining some insight into the general principles of UG that govern the mental computations underlying the use of language. [...] The more abstract are the principles, the more deeply embedded in a particular theoretical structure and remote from presented phenomena, the more interesting and significant is the study of language. (*ibidem*)

Regarding the search for those abstract principles which could be ascribed to UG, it must be pointed out that Chomsky had already begun this investigation in his early works. However, it was only in the Seventies that Chomsky finally managed to formulate a series of conditions on grammar rules – an issue that, until the late Sixties, had been dealt with

only from a programmatic perspective (cf. Graffi & Rizzi 1979, pp. 379-380). These conceptual advancements introduced in the EST, while making explicit some of the ideas which govern the development of Generative Grammar to this day, constitute the foundations on which Chomsky would base the theoretical framework known as the P&P model.

### 1.2.2 – The term “parameter” in *Conditions on Rules of Grammar* (1976)

Before looking at the point when the concept of parameter was officially and definitely born, it can be helpful to trace back the first appearance of the term “parameter” in Generative Grammar and to relate it to the context where it was originally used. This particular notion, although without the more specific meaning it would bear a few years later, first occurs apparently in Chomsky’s *Conditions on Rules of Grammar* (1976). This article can be considered as the natural development of the considerations originally expressed in *Conditions on Transformations* (1973), the work in which Chomsky laid out the line of research which characterized the EST phase of his linguistic research until the foundation of the P&P model (cf. Graffi & Rizzi 1979, p. 380).

At the beginning of *Conditions on Rules of Grammar*, Chomsky briefly brings back up the main issues and ideas which summed up the research program of Generative Grammar’s EST. While the theoretical assumptions mentioned here are exactly those originally presented during the EST phase, the objective proposed for the upcoming research, and which Chomsky hopes for, is far more focused than how it had been in the standard model’s phase. Here, immediate priority is given to finding some ways «to constrain the functioning of grammatical rules and thereby to limit the generative power of grammars of a given form» as anticipated in *Conditions on Transformations* (Chomsky 1977b [1973], p. 84):

A person who has learned a language has constructed a system of rules and principles – a grammar – determining a sound-meaning relation of some sort over an infinite domain. The linguist’s grammar is a theory of this attained competence, under conventional and entirely appropriate idealizations. The general theory of grammar – call it “universal grammar” (UG) – is a system of principles that determines: (1) what counts as a grammar and (2) how grammars function to generate structural descriptions of sentences. Thus within UG we have conditions on the form of grammar and conditions on the function of grammatical rules. [...]

Naturally, we will attempt to make these conditions explicit and as restrictive as possible.  
(Chomsky 1977b [1976], p. 163)

However, although in Chomsky's previous works there was no explicit reference to the possibility for primary linguistic data to actively exert an influence over the restrictive properties of UG, in this paper this scenario is not completely excluded. On the one hand, the universal conditions strictly inherent to UG correspond to «those properties of attained linguistic competence that hold by necessity rather than a result of accidental experience, where by necessity» Chomsky means «biological rather than logical necessity» (Chomsky 1977b [1976], p. 164). On the other hand, a certain role, although quite limited, in the development of the structure itself of UG is attributed by Chomsky to the speaker's own linguistic experience:

We can explain the fact that linguistic competence has the property *P* insofar as we can show that property *P* conforms to UG and is, furthermore, the special case of UG determined by experience. In most interesting cases, the role of experience is limited or even nonexistent so that the property *P* simply reflects some property of UG and thus gives us direct insight into the nature of UG. We argue that a given language has the property *P* because UG requires that this be the case. Where it seems that speakers have been exposed to little if any relevant experience, but yet have acquired a language with the property *P* rather than some alternative, it is reasonable to attribute *P* to UG itself. (*ibidem*)

This stance on the influence of experience over UG seems to indirectly anticipate the concept of parameter or, at least, to not contradict the conceptual scope it would be given from the early Eighties onward. According to Chomsky, the feasibility for experience to have a role in further restricting the set of conditions on rules provided by UG can be justified by the possible existence of conditions which, although «in the best case, of course, [...] will be universal», still «may be language-particular or rule particular» (Chomsky 1977b [1976], p. 175). Based on such conditions, what Chomsky refers to by using the term “parameter” in this specific context is simply a condition which is not universal, but is eligible to apply to a finite set of languages or grammatical rules:

Even if conditions are language- or rule-particular, there are limits to the possible diversity of grammar. Thus, such conditions can be regarded as parameters that have to be fixed (for the language, or for particular rules, in the worst case), in language learning. [...] It has often been supposed that conditions on application of rules must be quite general, even universal, to be significant, but that need not to be the case if establishing a “parametric” condition

permits us to reduce substantially the class of possible rules. (Chomsky 1977b [1976], p. 175)

This new approach to linguistic difference is still desirable as long as it contributes to the fulfilment of Generative Grammar's primary aim, that is, «to restrict the class of grammars available in principle to the language learner» without giving up universality (Chomsky 1977b [1976], p. 167). Moreover, Chomsky does not exclude that «we might hope to find that even if some condition C on rule application is language-particular, nevertheless some general principle determines that it applies in languages of some specific type» (Chomsky 1977b [1976], p. 175). In this case, the application of these language- or rule-particular conditions would actually be determined by a more general principle, thus not letting linguistic theory neither draw back on explanatory adequacy nor contradict the universality and innateness of language faculty:

Such a result would be as welcome as a universal condition, in that it limits the choice of grammars in a comparable way. That is, the child would not have to learn anything about the applicability of the condition; a universal principle would determine this. (*ibidem*)

Even if in the quotation proposed above the term “parameter” can potentially suggest the more complex meaning it bears in *Lectures on Government and Binding* (1981a), in this context this notion *per se* did not imply the structural relations it would be given in the P&P model yet. One important aspect which is worth pointing out is that, contrary to the general “parametric inquiry” which would shortly be conducted by linguists such as Rizzi, Taraldsen, Hale and Graffi, Chomsky did not explicitly dedicate his efforts to parametric variation among languages. On the one hand, Chomsky seemed to leave the door open for non-universal variation while, on the other, he kept focusing his attention exclusively on the principles constituting UG. This theoretical stance probably derived from the fact that, although at that time Chomsky's main research interests were revolving around principles which were strictly universal, he was nonetheless still eager to welcome any further perspective which could help Generative Grammar limit the set of possible grammars.



### 1.3 – Generative Grammar's development in the second half of the Seventies and the first explicit parameters

By the late Seventies, the search for linguistic universals yielded some very important contributions to the theory of Generative Grammar. Among all the programmatic developments proposed by Chomsky pointed out in the previous chapter, the well established purpose of constraining the form and the applicability of rules effectively led to the formulation of a set of general conditions which seemed good candidates for the role of principles of UG. Two of these potential principles, which in turn would have a huge impact on the very first researches about parametric variation, were “Subjacency” and “Nominative Island Condition” (NIC). In order to better introduce the first parameters explicitly proposed in Generative Grammar, before looking at the works of Rizzi (1978) and Taraldsen (1978) there will be a brief presentation of these two universal conditions which, in turn, would respectively be the core of these seminal papers.

#### *1.3.1 – The Subjacency Condition and the Nominative Island Condition*

Although sharing the same general function of limiting the set of possible grammars, while the Subjacency Condition was conceived as a condition on movement rules, limiting the possible range of a single syntactic movement, the NIC was conceived as a condition on representations, as it accounted for the distribution of pronominal anaphora and the syntactic relationship between these elements and their antecedents (cf. Graffi 2001, p. 449).

The first of these conditions to be presented here is the Subjacency Condition. Although Subjacency was first postulated in Chomsky (1973), the formulation proposed here is taken from Chomsky (1977a) because of its greater clarity:

I will understand the subjacency condition as holding that a cyclic rule cannot move a phrase from position Y to position X (or conversely) in:

...X...[<sub>α</sub>...[<sub>β</sub>...Y...]....]...X..., where α and β are cyclic nodes.

(Chomsky 1977a, p. 73)

The notions of “cyclic rule” and “cyclic node” mentioned in this definition refer to a peculiar aspect of movement rules known as Chomsky’s principle of the “cycle”, which was first discussed in Chomsky (1965). This principle states that rules belonging to the specific type of transformational rules – namely, according to the terminology of the late EST phase, the rules which can be «restricted to the single rule: Move  $\alpha$ , where  $\alpha$  is a category» (Chomsky 1980, p. 3) – must apply «sequentially, "from the bottom up" – that is, applying the sequence of rules to a given configuration only if we have already applied it to all base Phrase-markers embedded in this configuration» (Chomsky 1965, p. 143). In these terms, movement rules are cyclic since, after applying to the most embedded clause of a complex sentence, they can apply to the clause immediately dominating it – and so on – only on condition that the former does not require any further transformation. As Chomsky writes:

[...] the grammar contains a linear sequence of singular transformations. These apply to generalized Phrase-markers cyclically, in the following manner. First, the sequence of transformational rules applies to the most deeply embedded base Phrase-marker. [...] Having applied to all such base Phrase-markers, the sequence of rules reapplies to a configuration dominated by S in which these base Phrase-markers are embedded [...], and so on, until finally the sequence of rules applies to the configuration dominated by the initial symbol S of the entire generalized Phrase-marker [...]. That is, singular transformations are applied to constituent sentences before they are embedded, and to matrix sentences after embedding has taken place. (Chomsky 1965, pp. 134-135)

With regards to the notion of “cyclic node”, according to Chomsky’s analysis of English grammar «within the extended standard theory [...] both NP and S are nodes to which cyclic operations apply» (Chomsky 1977b [1973], p. 91). Based on such conditions, Subjacency accounts for the agrammaticality of those sentences, such as (1), which would derive from a movement rule crossing two cyclic nodes at a time (cf. Chomsky 1977b [1973], p. 103):

- (1) \*Who does he believe the claim that John saw?
- (2) He believes [<sub>NP</sub> the claim [<sub>S</sub> John saw who]]

In this case, the rule that would derive (1) would move the wh-element “who” from its base position in (2) to the left periphery of the sentence crossing both a NP node and a S node in a single movement. Hence Subjacency effectively rules out those sentences in

which a syntactic movement involves two cycles of which the former «is not subjacent» to the latter (Chomsky 1977b [1973], p. 103).

Now it is the turn of NIC. Originally postulated as “Tensed Sentence Condition” (TSC) in Chomsky (1973), and after undergoing progressive revision, the condition on representations known as NIC received its definitive formulation in the essay *On Binding* (Chomsky 1980), where it was defined as follows:

- (3) A nominative anaphor cannot be free in S'.

(Chomsky 1980, p. 36)

This definition requires some further clarification. According to Chomsky, while «lexical NPs are not anaphors», «PRO, trace, and reciprocal are anaphors» (Chomsky 1980, p. 10). Leaving reciprocals aside, by the terms “PRO” and “trace” Chomsky refers to two distinct kinds of *empty categories*, a notion which in turn identifies syntactic categories without phonological content:

Assuming the notation of labeled bracketing, we stipulate the following base convention: if the category  $\alpha$  is not expanded in a derivation, then apply the rule [4], where  $e$  is the identity element:

- (4)  $\alpha \rightarrow [\alpha \ e]$

(Chomsky 1980, pp. 3-4)

On the one hand, Chomsky generally identifies as traces those empty categories – conventionally marked by the symbol  $t$  – which result when a movement rule moves a syntactic category – in the particular case addressed below, a NP – from its base position into another position:

Movement of the category  $\alpha$  is assumed to “leave behind” the category  $[\alpha \ e]$ , in accordance with trace theory. [...] Assume that  $\alpha$  and its trace are coindexed, by convention. (Chomsky 1980, p. 4)

The “coindexing” referred to in the above quotation is the conventional way to mark coreference between an anaphor (or a pronoun) and its antecedent. In fact, it should be pointed out that a trace does not merely indicate the original position left by a rule of NP-movement but, although not being lexical, it still «plays a role in determining the well-

formedness of the sentence» by behaving like an NP without phonetic realization (Graffi 2001, pp. 432-433). On the other hand, according to Chomsky «we may take PRO to be just base generated  $t(x)$ ,  $x$  a variable» (Chomsky 1977a, p. 82). In other words, the label PRO indicates the specific kind of empty category introduced in a base derivation – hence not left by a rule of movement – which corresponds to «a NP without a fixed index», and whose index «is then assigned by a rule of control» in such sentences as (5) (*ibidem*):

(5) I persuaded the man<sub>i</sub> [<sub>S</sub> PRO<sub>i</sub> to leave]

A structure as (5) is said to be a control structure, since the embedded NP which corresponds to the subject of the infinitival subordinate clause is controlled by the subject of the matrix clause and thus it receives its index:

The element [<sub>NP</sub>  $e$ ] [...] is what is conventionally represented PRO, a position that must undergo control. (Chomsky & Lasnik 1977, p. 432)

Summing up, according to Chomsky «it follows, then, that trace and PRO are the same element», as both are instances of empty categories, and that «they differ only in the way the index is assigned – as a residue of a movement rule in one case, and by a rule of control in the other» (Chomsky 1977a, p. 82).

Finally, the term “free” refers to the case where an anaphor  $\alpha$  has no possible coreferential element in the syntactic domain  $\beta$ . Therefore, NIC necessarily requires that, if  $\alpha$  is an anaphor in the domain of the tensed sentence and, at the same time, is assigned nominative Case for being the subject of the clause, then  $\alpha$  must have an antecedent within the domain of the tensed sentence; otherwise, the resulting sentence is ruled out. Based on such conditions, a sentence such as (6) would be ruled out because the only possible antecedent of the nominative anaphor lies outside S' (Chomsky 1980, p. 14):

(6) They expected [<sub>S</sub> that each other would be there]

In fact, the antecedent “they” does not occur within the tensed sentence whose subject corresponds to the reciprocal “each other”, hence the latter has no possible coreferential element within S': it is free or, conversely speaking, it is *not bound* in S'. NIC is therefore a condition on representations, since in cases such as (6) it rules out the coreferential interpretation between nominative anaphors and their antecedents.

### *1.3.2 – From a language-particular approach to a comparative approach in the investigation of UG*

One of the central claims of Generative Grammar as a theory of language, and which is also one of the implications of the innateness of language faculty, is that, since the principles of UG exert their influence over all possible languages without any exception, «a great deal can be learned about UG from the study of a single language, if such study achieves sufficient depth to put forth rules or principles that have explanatory force but are undetermined by evidence available to the language learner» (Chomsky 1993 [1981a], p. 6). Being Generative Grammar based on a deductive approach, from this perspective it is entirely reasonable to expect that «the principles that appear to have explanatory adequacy for English are the principles of universal grammar» (Chomsky 1975b, p. 118). Looking at the theoretical developments brought forth during the standard model and the EST phase, this logical assumption was the very basis of Chomsky's own quest for universal principles since the early systematization of Generative Grammar:

Study of a wide range of languages is only one of the ways to evaluate the hypothesis that some formal condition is a linguistic universal. Paradoxical as this may seem at first glance, considerations internal to a single language may provide significant support for the conclusion that some formal property should be attributed not to the theory of the particular language in question (its grammar) but rather to the general linguistic theory on which the particular grammar is based. (Chomsky 1965, p. 209)

In fact, although his linguistic analysis was based solely on English grammar, it effectively led to the formulation of some very general conditions which, as Subadjacency and NIC, had all what it takes to be assumed as universal. However, while «it is important to bear in mind that the study of one language may provide crucial evidence concerning the structure of some other language, if we continue to accept [...] that the capacity to acquire language [...] is common across the species» (Chomsky 1986b, p. 37), the empirical contribute from languages other than English soon proved to be essential to the development of Chomsky's theory of language. As Generative Grammar proceeded and its approach to language analysis spread outside the United States, more and more linguists whose native language was not English contributed the analysis of their respective mother tongue to Chomsky's line of inquiry. It is in this very scenario that Generative Grammar, after two decades from its birth, took an unexpected turn from being

focused on one language towards the systematic analysis of typological differences between distinct languages.

### 1.3.3 – The first example of parametrization: Rizzi and the principle of Subjacency

As stated by Baker, «a discipline really gets interesting when the ideas of the thinkers begin to converge with the empirical discoveries of the doers» (Baker 2001, p. 35). In the case of Generative Grammar, the convergence between the deductive hypotheses on UG's architecture based on the analysis of English grammar and empirical evidence from other languages first happened as a result of an observation made by the Italian linguist Luigi Rizzi in the late Seventies. In the essay *Violations of the Wh island constraint in Italian and the subjacency condition*, published in 1982 but «written in the fall of 1977 during a stay at MIT» (Rizzi 1982, p. xii), Rizzi observed that Italian allows sentences as the following (Rizzi 1982 [1978], p. 50):

- (7) Il solo incarico [<sub>S'</sub> che<sub>i</sub> [<sub>S</sub> non sapevi [<sub>S'</sub> a chi<sub>j</sub> [<sub>S</sub> avrebbero affidato t<sub>i</sub> t<sub>j</sub> ]]]] è poi finito proprio a te.
- (8) Tuo fratello, [<sub>S'</sub> a cui<sub>i</sub> [<sub>S</sub> mi domando [<sub>S</sub> che storie<sub>j</sub> [<sub>S</sub> abbiano raccontato t<sub>i</sub> t<sub>j</sub> ]]]], era molto preoccupato.
- (9) La nuova idea di Giorgio, [<sub>S'</sub> di cui<sub>i</sub> [<sub>S</sub> immagino [<sub>S'</sub> che cosa<sub>j</sub> [<sub>S</sub> pensi t<sub>i</sub> t<sub>j</sub> ]]]], diverrà presto di pubblico dominio.

The problem represented by the well-formedness of these Italian sentences is that, in principle, this fact should constitute a violation of Subjacency, a condition which until then had all the requisites for representing an actual property of UG, thus falsifying Chomsky's hypothesis. For example, in (7) the relative pronoun “che<sub>i</sub>” can be extracted from the embedded indirect question and, from its base position t<sub>i</sub>, it can cross two S nodes and reach its destination, COMP<sub>2</sub>, necessarily by means of a single wh-movement, since the COMP<sub>1</sub> position – namely, its only possible «“escape hatch”» in Rizzi's words – is already filled by the wh-clause “a chi<sub>j</sub>” (*ibidem*). However, the efficiency of Subjacency does not seem to be called into question in English, since the respective translations of sentences (7), (8) and (9) are not grammatical, as shown by (10), (11) and (12) (*ibidem*):

- (10) \*The only charge [<sub>S'</sub> that<sub>i</sub> [<sub>S</sub> you didn't know [<sub>S'</sub> to whom<sub>j</sub> [<sub>S</sub> they would entrust t<sub>i</sub> t<sub>j</sub> ]]]], has been entrusted exactly to you.
- (11) \*Your brother, [<sub>S'</sub> to whom<sub>i</sub> [<sub>S</sub> I wonder [<sub>S'</sub> which stories<sub>j</sub> [<sub>S</sub> they told t<sub>i</sub> t<sub>j</sub> ]]]], was very troubled.
- (12) \*Giorgio's new idea, [<sub>S'</sub> of which<sub>i</sub> [<sub>S</sub> I imagine [<sub>S'</sub> what<sub>j</sub> [<sub>S</sub> you think t<sub>i</sub> t<sub>j</sub> ]]]], will soon become known to everybody.

At this point, assuming that the «bounding nodes» referred to by Subjacency be NP and S (Rizzi 1982 [1978], p. 49), according to Rizzi «the derivation of sentences [(7), (8) and (9)] is possible only at the cost of violating some conditions on rules, or significantly changing the syntax of the complementizer» (Rizzi 1982 [1978], p. 52). More specifically, a sentence like (9) could only be derived (*ibidem*):

- (13) (A) By violating the strict cyclicity condition: *di cui* is moved into COMP1 at the cycle S'1, then to COMP2 at the cycle S'2, and then *che cosa* is moved into COMP1.
- (B) By allowing a COMP to be filled more than once per cycle: *che cosa* and *di cui* are both moved to COMP1 at the first cycle, and that *di cui* is moved to COMP2 at the second cycle.

However, Rizzi's ingenious solution to this apparently unsolvable problem was to suggest that, contrary to appearances, Subjacency still holds, and that the reason is that the notion of cyclic node varies from language to language. More precisely, Rizzi proposed that the subjacency principle is *parametrized* – using today's terminology. To put it simply, although Subjacency is a universally operative principle, «different languages have different bounding nodes» (Newmeyer 2005, p. 39). With regards to the distinction between S and S' formulated in Bresnan (1970), according to which S is constituted by a subject and a predicate, while S' is constituted by a complementizer and a phase S meant as a «subject-predicate “nucleus”» (Graffi 2001, p. 431), according to Rizzi in Italian «the bounding node which is relevant for subjacency is S', not S», hence «wh-movement does obey the Subjacency Condition» without being ruled out (Rizzi 1982 [1978], p. 57). This very elegant solution explains the reason why the Italian sentences proposed in the examples above are well-formed, while their respective English translations would not be possible due to their violation of Subjacency.

### 1.3.4 – Taraldsen and the Null Subject parameter

Although Rizzi (1982) was the first work to explicitly attribute a typological variation to a different parametric setting, it was not the only attempt made in the late Seventies to relate the apparent violation of an universal constraint to a specific parametric configuration. In this sense, in those years there was also another linguist whose idea actively contributed to the shaping of the P&P model. This is the case of Tarald Taraldsen, a Norwegian linguist who, in his essay *On the NIC, vacuous application and the that-trace filter* (1978), postulated that there is «a difference in the behaviour of empty categories across languages» (Graffi 2001, p. 448). Taraldsen's proposal aimed at accounting for the fact that some languages seemed to be exempt from NIC; a condition which, as Subjacency, was commonly regarded as a universal constraint. In his paper, Taraldsen observed that «the NIC is inoperative [...] with null anaphors in Italian» in at least two respects: first, Italian allows silent definite pronominal subjects without incurring a violation of NIC; second, traces created by wh-movement in Italian are not ruled out by NIC (Taraldsen 1978, p. 2):

The first of the two arguments to be presented here shows that a relatively plausible analysis of resumptive pronouns in Italian implies that some occurrences of null subjects are occurrences of [<sub>NP</sub> e] that are not affected by the NIC. The second argument shows that traces created by wh-movement also are insensitive to the NIC in the same way. (*ibidem*)

Therefore, the problem addressed by Taraldsen is that the occurrence of null (i.e., *empty*) subjects and nominative traces in Italian, both being instances of nominative anaphors, seemed to falsify NIC. For example, while «both Italian and French permit extraction from embedded interrogatives [...] only Italian, however, permits the extraction of the subject of an embedded tensed interrogative» as in sentences (14) and (16) (Taraldsen 1978, pp. 5-6):

- (14) Abbiamo spedito questionari a chi non sapevamo per chi simpatizzasse.

We have sent questionnaires to whom we didn't know for whom were sympathizing

- (15) \*Nous avons envoyé des questionnaires à qui nous ne savions pas avec qui sympathisait.



- (16) Inviteremo tante ragazze quante ci ricordiamo ancora dove abitano.  
We will invite as many girls how many we still remember where are living

With regards to traces created by wh-movement, in order to propose an alternative to this apparent invalidation of NIC, Taraldsen first discusses the hypothesis that such sentences as (14) and (16) could be derived «by deletion of unstressed subject pronouns» (Taraldsen 1978, p. 5). If this assumption were true, these anaphors would actually be “gaps” corresponding to previously deleted pronouns, thus not triggering NIC’s effect at all (cf. Taraldsen 1978, p. 6). According to Taraldsen, however, since «in general, the fronted wh-phrases in these examples do not co-occur with resumptive pronouns wherever these would be visible» (*ibidem*), as shown in the ungrammaticality of sentences such as (17) (Taraldsen 1978, p. 3):

- (17) \*Ecco la ragazza che sembra che lei voglia bene a Giorgio.  
Here is the girl that it seems that she loves George

this does not seem a viable option:

Independently of the claims in the preceding paragraphs, [14]-[16] are not good candidates for an analysis in terms of resumptive pronouns that would claim that the relevant +nom gaps contain pronouns at the point of application of the NIC. (Taraldsen 1978, p. 6)

Excluding the possibility that (14) or (16) could be analyzed in terms of resumptive pronouns, according to Taraldsen it is clear that «the NIC is inoperative with traces of wh-phrases in Italian» (*ibidem*). Moreover, as anticipated above, also null subjects in Italian seem to be exceptional with respect to the NIC. This is clear in sentences as (18), whose silent subject corresponds to the empty NP in (19), yet does not lead to ill-formedness (Taraldsen 1978, p. 11):

- (18) Sono troppo pigri.  
They are too lazy  
(19) [<sub>NPi</sub> e] sono<sub>i</sub> troppo pigri

Based on such conditions, Taraldsen assumed the existence, in languages such as Italian, of some special device capable of governing nominative anaphors which would not

be active in languages such as French. This intuition, which become known as “Taraldsen’s generalization”, corresponds to the idea that there is a connection between the occurrence of empty subjects in finite clauses and the morphological “richness” of the verbal agreement. More precisely, according to Taraldsen the NIC does not apply to null subjects in Italian because the agreement feature of the finite verb binds the nominative anaphor:

Our approach will be the following: rather than readjust the NIC in some unknown way or simply take the Italian exceptionality to the NIC as a primitive, we will analyze Italian in such a way that [+nom] occurrences of  $[_{NP} e]$  are not free (i) in their minimal  $S'$ . Drawing on a traditional insight, we take the subject-verb agreement in Italian and other “null subject languages” to be the decisive factor. Formally, we will represent a subject NP as being co-indexed with a finite verb and assume that it for that reason is bound in “its”  $S'$ . (Taraldsen 1978, p. 7)

On this approach, the relevant parts of sentences (14) and (16) are represented as (20) and (21) (*ibidem*):

(20)  $[S' \text{ per chi } t_i \text{ simpatizzasse}_i]$

(21)  $[S' \text{ dove } t_i \text{ abitano}_i]$

and, since the traces left by subject extraction are not free in  $S'$ , there is absolutely no violation of the NIC. In addition to this, «by analyzing null subjects as empty NPs rather than using a rule of pronoun deletion» the validity of the NIC is not called into question anymore by sentences such as (22) – which represents the revised version of (19) (Taraldsen 1978, p. 11):

(22)  $[S' [_{NP}_i e] \text{ sono}_i \text{ troppo pigri}]$

In these terms, the basic factor determining the distinction between «pro-drop» (i.e., null subject languages) and «non-pro-drop languages» can be identified with the parametrization of verb inflection (Chomsky 1993 [1981a], p. 240). According to Rizzi, who after Chomsky (1981a) further developed Taraldsen’s intuition, the NIC does not interfere with null anaphors in languages such as Italian because pro-drop languages allow the inflection to be optionally specified [+ pronoun]. If so specified, «a verbal affix is interpreted

as a definite pronoun, and permits an argument NP position to be empty» (Rizzi 1982, p. 130):

[...] the characteristic property of NSL's is that their verbal inflections have (clitic-like) pronominal properties. This intuition can be straightforwardly implemented by assuming that INFL in NSL's is specified with the feature [+ pronoun]: i.e., like a clitic, it is a verbal affix with (pro-) nominal properties, specified with respect to such grammatical features as person and number; and, like a clitic, it is interpreted as a definite pronoun [...], and binds and properly governs an empty NP position. (Rizzi 1982, pp. 130-131)

On the other hand, while «we accordingly predict that a language can have “null subjects” in the same way as Italian [...] only if a [+nom] occurrence of [NP<sub>i</sub> e] escapes the NIC in a comparable way» (Taraldsen 1978, p. 12), in languages such as French the verbal inflection is obligatory specified [– pronoun], and therefore any nominal anaphor free in S' would be ruled out. On par of Rizzi (1982), Taraldsen (1978) successfully showed that the violation of a putatively universal constraint is only apparent, and that the distinct behaviour of empty categories between languages such as Italian and French can be attributed to a different parametric setting.

### *1.3.5 – Toward the systematic formulation of the P&P model*

The implications of Rizzi's intuition that small changes in the parameters would lead to major effects on generated structures propelled Generative Grammar into an entirely new phase. From that moment onward, the dominating approach characterizing this theoretical framework was no longer only «to abstract statements and generalizations from particular descriptively adequate grammars and, wherever possible, to attribute them to the general theory of linguistic structure» (Chomsky 1965, p. 46). On the other hand, the research program of the P&P model attempted «to attribute differences among languages to slightly different settings of the parameters associated with [...] principles» provided by our innate language faculty (Newmeyer 2005, p. 41).

From a theoretical point of view, both Rizzi and Taraldsen's works had the crucial – and, in that moment, unprecedented – effect of showing that there can actually be a strong correlation between two otherwise independent linguistic phenomena like, on the one hand, the possibility of having null subjects and, on the other, the possibility of having wh-movement of a subject through an explicit complementizer (cf. Taraldsen 1978, p. 2). This

sort of empirical confirmations, together with the fact that previously unexplainable typological differences between languages could now be attributed to minor parametric changes, triggered an enormously productive research program. In fact, after the transition from the EST to the GB-theory, linguistic variation would be no more regarded only as a testing ground for UG hypotheses, but it became an actual means by which potentially derive all the facets of linguistic complexity.

#### 1.4 – The concept of parameter as an epistemological necessity

As the number of languages whose grammars were taken into consideration appeared to increase, Generative Grammar could no more exempt itself from taking into account the advantage of approaching UG's architecture also from the side of linguistic variation rather than only «from the study of a single language» (Chomsky 1993 [1981a], p. 6). Chomsky himself, whose own research has mainly focused on English grammar and the discovery of strictly universal conditions, became more and more interested in the fact that «language seemed to systematically vary with respect to certain features of Universal Grammar» (Graffi 2001, p. 449). This interest, which had already been showed in Chomsky (1976), thanks to Rizzi and Taraldsen's contributions kept growing until the notions of principle and parameter were finally incorporated into a systematic and coherent theoretical model.

##### *1.4.1 – The concept of parameter in the “Pisa Lectures”*

After the establishment of the P&P model in the early Eighties, Generative Grammar underwent a major change in its attitude toward cross-linguistic variation. The reference work of this model, which is represented by Chomsky's *Lectures on Government and Binding* (1981a), is based on lectures which Chomsky gave at the GLOW conference and workshop held at the Scuola Normale Superiore of Pisa in April 1979 (cf. Chomsky 1993 [1981a], p. vii). While the central aim of the theory proposed in the so-called “Pisa Lectures” was to account for the phenomena of binding across languages, the P&P approach carried out in this work was at the same time «the first real effort made within the Chomskian program to provide a systematic account of cross-linguistic differences» (Graffi 2001, p. 454).

In Chomsky (1981a), the new approach to linguistic variation which would be carried out from the GB-Theory's phase onward is outlined in the first chapter, which is entitled *Outline of the Theory of Core Grammar*. In its very beginning, the first aspect to be pointed out by Chomsky is Generative Grammar's compelling need to develop a linguistic theory which is both highly structured and open to variation. In order to really meet the level of explanatory adequacy, a theory of UG must necessarily be highly structured enough to provide a limit to the number of possible grammars. At the same time, however, its architecture must be sufficiently open to allow for language variation:

Let us recall the basic character of the problem we face. The theory of UG must meet two obvious conditions. On the one hand, it must be compatible with the diversity of existing (indeed, possible) grammars. At the same time, UG must be sufficiently constrained and restrictive in the options it permits so as to account for the fact that each of these grammars develops in the mind on the basis of quite limited evidence. (Chomsky 1993 [1981a], p. 3)

These theoretical requirements, which are not dissimilar from the ones previously outlined in the *Aspects*, are nonetheless complemented by an additional property which clearly refers to the approach first implemented by works such as Rizzi (1982) and Taraldsen (1978). Together with explanatory adequacy, another hypothetical prerequisite of a desirable linguistic theory would be its ability to account for a broad range of empirical phenomena by means of slight changes individually operated to the whole architecture of UG:

In work of the past several years, [...] several theories have been proposed that are fairly intricate in their internal structure, so that when a small change is introduced there are often consequences throughout this range of phenomena, not to speak of others. This property of the theories I will investigate is a desirable one; there is good reason to suppose that the correct theory of universal grammar in the sense of this discussion (henceforth: UG) will be of this sort. (*ibidem*)

By means of such a theory, Generative Grammar would in fact be able not only to reconcile language universality and typological variation, but also to account for previously unexplainable clustering of grammatical properties. Moreover, any discovery in this specific direction would in turn be evidence that this sort of knowledge is acquired on a more abstract basis than what can be inferred or taught from the outside, thus further supporting the explanatory adequacy of this new theoretical approach. In these terms, the

central idea of Chomsky's hypothesis is that all possible languages are built on a common core of structural *principles* which exert their systematic influence on grammatical structures and rules without any exception, and that linguistic variation is due to a limited number of *parameters*: a set of universal choices, each with a finite number of values, left unspecified by UG and whose setting depends on the experience of native speakers during language acquisition:

What we expect to find, then, is a highly structured theory of UG based on a number of fundamental principles that sharply restrict the class of attainable grammars and narrowly constrain their form, but with parameters that have to be fixed by experience. If these parameters are embedded in a theory of UG that is sufficiently rich in structure, then the languages that are determined by fixing their values one way or another will appear to be quite diverse, since the consequences of one set of choices may be very different from the consequences of another set; yet at the same time, limited evidence, just sufficient to fix the parameters of UG, will determine a grammar that may be very intricate and will in general lack grounding in experience in the sense of an inductive basis. (Chomsky 1993 [1981a], pp. 3-4)

Although in Chomsky (1976) it had already been suggested that linguistic complexity could possibly be limited by additional conditions of UG which are «language-particular or rule particular» instead of strictly universal (Chomsky 1977b [1976], p. 175), it is only in Chomsky (1981a) that the concept of parameter assumes its definitive meaning. In the former work, which represented the EST phase, parameters were regarded as secondary «language- or rule-particular» conditions by which to restrict the number of possible grammars (*ibidem*). In the P&P model, however, because of its power to predict the clustering of typological properties, the concept of parameter becomes as important as that of principle. In fact, the limiting power of the P&P model derives not only from the restricted number of the parametric options allowed by UG, but also from the fact that, according to this theory, there are likely to be structural relations between parameters and their settings in such a way that an implicational hierarchy among them can be posited:

In a highly idealized picture of language acquisition, UG is taken to be a characterization of the child's pre-linguistic initial state. Experience – in part, a construct based on internal state given or already attained – serves to fix the parameters of UG, providing a core grammar, guided perhaps by a structure of preferences and implicational relations among the parameters of the core theory. (Chomsky 1993 [1981a], p. 7)

This implicational organisation among parameters would explain the correlation between two or more otherwise independent grammatical objects or structures occurring in the same language, and thus the reason why certain typological features tend to co-occur in clusters. Moreover, this “directional” nature of parametric setting can also potentially account for the order of first language acquisition by providing a necessary sequence in which parameter need to be fixed, thus generating increasingly complex grammatical stages.

#### 1.4.2 – Language typology in the P&P model: core grammar and periphery

Besides the distinction between principles and parameters, one of the key concepts of the P&P approach is that of *core grammar*. By this notion, Chomsky indicates any permitted combinations of values for any parameter and, at the same time, any grammar whose structure is determined exclusively by a possible parametric setting:

When the parameters of UG are fixed in one of the permitted ways, a particular grammar is determined, what I will call a “core grammar”. (Chomsky 1993 [1981a], p. 7)

Although this concept has a central role in Chomsky (1981a), it is worth pointing out that its introduction in Generative Grammar dates back to some years earlier, and more precisely to Chomsky & Lasnik’s *Filters and Control* (1977). In this work, the notion of core grammar is closely related to that of *markedness*, which it is meant in this context both as a set of restrictions exerting their pressure towards the selection of certain structural configurations over others and an evaluation measure whose unmarked cases are represented by “core” processes:

We will assume that UG is not an “undifferentiated” system, but rather incorporates something analogous to a “theory of markedness”. Specifically, there is a theory of core grammar with highly restricted options, limited expressive power, and a few parameters. Systems that fall within core grammar constitute “the unmarked case”; we may think of them as optimal in terms of the evaluation metric. (Chomsky & Lasnik 1977, p. 430)

In addition to the optimal and unmarked core processes, according to Chomsky & Lasnik «an actual language is determined by fixing the parameters of core grammar and then adding rules or rule conditions, using much richer resources» (*ibidem*). From this

perspective, non-core processes, which «we may think of as the syntactic analogue of irregular verbs», represent those «added properties of grammars» not directly captured by UG (Chomsky & Lasnik 1977, p. 430). Although Chomsky & Lasnik (1977) surely represents a development of the concept of parameter with respect to Chomsky (1976), in this work no internal distinction is made within the domain of core grammar, neither in terms of markedness nor concerning the possible implicational relations between parameters. On the other hand, in the P&P framework of Chomsky (1981a) the set of all core grammars would correspond not only to every typological generalization observable across different languages, but also to the finite «set of grammars resulting from the fixing of the parameters of UG in all possible ways», whether their settings are marked or unmarked (Newmeyer 2005, p. 47). As pointed out in both Chomsky & Lasnik (1977) and Chomsky (1981a), however, core grammar's scope cannot embrace the totality of the existing languages alone; in fact, in order to denote all observable grammars, the P&P approach has also to rely on another concept: that of *periphery*:

But it is hardly to be expected that what are called "languages" or "dialects" or even "idiolects" will conform precisely or perhaps even very closely to the systems determined by fixing the parameters of UG. This could only happen under idealized conditions that are never realized in fact in the real world of heterogeneous speech communities. Furthermore, each actual "language" will incorporate a periphery of borrowings, historical residues, inventions, and so on, which we can hardly expect to – and indeed would not want to – incorporate within a principled theory of UG. (Chomsky 1993 [1981a], pp. 7-8)

Outside the set of core grammars directly determined by UG, Chomsky points out that «what is actually represented in the mind of an individual even under the idealization of a homogeneous speech community would be [...] a periphery of marked elements and constructions» (Chomsky 1993 [1981a], p. 8). Although it somewhat represents a specific theoretical device postulated to account for all the idiosyncrasies which fall outside all logically possible parameters' values, according to Chomsky the notion of periphery does not simply consist in an arbitrary grey area destined to dismiss all grammatical irregularities. Rather, since «marked structures have to be learned on the basis of slender evidence too, [...] there should be further structure to the system outside of core grammar» (*ibidem*) such that «we do not expect to find chaos in the marked periphery of language» (Chomsky 1993 [1981a], p. 70):



We might expect that the structure of these further systems relates to the theory of core grammar by such devices as relaxing certain conditions of core grammar, processes of analogy in some sense to be made precise, and so on, though there will presumably be independent structure as well: hierarchies of accessibility, etc. (Chomsky 1993 [1981a], p. 8)

While in the EST phase every possible marked structure was only identified with non-core processes, according to the P&P model markedness is not exclusive to periphery, but it also occurs within core grammar. In these terms, Chomsky assumes two distinct manifestations of markedness in grammar: on the one hand, markedness within UG is represented by those parametric configurations which are selected when experience conflicts with the default and unmarked options referred to as «a structure of preferences [...] among the parameters» (Chomsky 1993 [1981a], p. 7); on the other hand, outside UG there are all those marked elements or structures resulting from the unavoidable interferences between Generative Grammar's idealized view of core grammar and all those factors which, both directly and indirectly, are extrinsic to UG (cf. Chomsky 1993 [1981a], p. 8).

#### *1.4.3 – The concept of parameter and language acquisition*

With regards to language acquisition, although Chomsky (1981a) does not propose any modifications to the solution to the problem of the poverty of the stimulus outlined in the *Aspects*, in the P&P model the concept of parameter is put side by side with the universal principles of UG in allowing the construction of the learners' linguistic competence. In addition to this, the preferential relations between parameters specified by markedness are assumed to be an additional device which integrates the child's pre-linguistic initial state in the task of language learning.

Returning to our idealized – but not unrealistic – theory of language acquisition, we assume that the child approaches the task equipped with UG and an associated theory of markedness that serves two functions: it imposes a preference structure on the parameters of UG, and it permits the extension of core grammar to a marked periphery. Experience is necessary to fix the values of parameters of core grammar. In the absence of evidence to the contrary, unmarked options are selected. (*ibidem*)

As the central aim of Generative Grammar remains that of achieving explanatory adequacy by shedding light on our biological endowment for language, the fact that instances of markedness can be found in both core grammar and periphery makes it difficult to distinguish markedness within core grammar from markedness within the periphery, that is, the marked parametric values of UG from all the other peripheral properties. However, since language acquisition is involved with core grammar and periphery in different ways, according to Chomsky «one would hope that evidence from language acquisition would be useful with regard to determining the nature of the boundary or the property of the distinction» (Chomsky 1993 [1981a], p. 9).

Finally, the most important advantage of the P&P theory with respect to language acquisition is that the architecture itself of this theoretical model works as an actual limit to the number of core grammars. That is, «UG will provide a finite set of parameters, each with a finite number of values» (Chomsky 1993 [1981a], p. 11) and, being this the only room for variation allowed within UG, the concept of parameter elegantly contributes to «the objective of reducing the class of grammars compatible with primary linguistic data» (Chomsky 1993 [1981a], p. 13):

It is worth asking whether the correct theory of UG does in fact permit only a finite number of core grammars. The theories that are being studied along the general lines I will be discussing here do have that property, and I think that it probably is the right property.  
(*ibidem*)

Because of the highly structured architecture of this theoretical model, while the «study of closely related languages that differ in some clustering of properties is particularly valuable for the opportunities it affords to identify and clarify parameters of UG» (Chomsky 1993 [1981a], p. 6), the search for parameters would in turn be particularly fruitful in the task of Generative Grammar to discover those universal principles which are assumed to characterize the language faculty and are determinant in accounting for language acquisition. In these terms, even if «it might be that this guiding intuition is mistaken» as language faculty, on par with other biological systems, may prove to «exhibit redundancies and other form of complexity» which would contradict the ideal simplicity of core grammar's principles (Chomsky 1993 [1981a], p. 14), according to Chomsky what really matters still is «unearthing a more “elegant” system of principles that achieves a measure of explanatory success» (Chomsky 1993 [1981a], p. 15).

## 1.5 – Conclusions

The history of the concept of parameter, from its theoretical premises to its definitive formulation, in a sense represents Generative Grammar's constant efforts to factor out the innate and hence universal properties of the biological system referred to as the faculty of language. Through all the steps involved in finding an abstract yet systematic way to account for linguistic variation, this concept was first mentioned from a purely hypothetical point of view (Chomsky 1976), and only after its empirical and more specific application in works such as Rizzi (1978) and Taraldsen (1978) it became an essential part of Chomsky's linguistic theory. However, although these "pre-parametric" inquiries brought forward some of the ideas which would be presented in Chomsky (1981a), on the other hand their main purpose was essentially to maintain a degree of descriptive adequacy which seemed to have been lost after the universal conditions formulated in the EST phase were called into question. Therefore, Rizzi and Taraldsen's works had surely the merit of triggering the transition from the EST to the P&P model, especially by their implication of allowing a single feature to account for a plurality of typological properties, but it was in Chomsky (1981a) that the concept of parameter definitely acquired those structural traits which would be important in achieving explanatory adequacy within a theory of linguistic acquisition.

With regards to an evaluation of the concept of parameter, the aspects highlighted in this chapter alone do not allow any judgement to be made. However, what seems to be self-evident is that, regardless of the effective success of the P&P model, right between the end of the EST phase and the advent of GB-Theory the parametric approach represented an epistemological necessity. In fact, in that precise moment, the concept of parameter was the ideal candidate for both further restricting the linguistic complexity allowed by UG principles and accounting for previously unexplainable typological generalizations occurring within language variation.



## **Chapter II**

### **The formulation of the main parameters of the Government-Binding Theory**

#### **2.1 – A retrospective classification of the parameters of the Government-Binding Theory**

Before attempting to delve into an analysis of their individual formulation, here I sketch one possible classification of the main parameters that were proposed in Generative Grammar during the Eighties. This classification is in turn built on the list of parameters that is proposed in Rizzi (2014), which has strengths in both its heterogeneity and theoretical value. Although the technical development of Generative Grammar has seen some of these parameters being abandoned at a later time, the range and diversity of the linguistic phenomena they were conceived to account for at the time of their formulation is presented as a proof not only of the pervasivity of the concept of parameter itself, but also of the role this notion has played in the search for explanatory adequacy pursued by generative linguists.

The parameters included in the following list span through such distinct domains as locality, government, binding and abstract Case, not to mention properties of syntactic movement, subcategorization and linearization. The main criterion adopted in the classification proposed here has been to group parameters according to the specific syntactic property they would refer to in current minimalist theories. It goes without saying that some of the theoretical devices originally referred to by these parameters have in turn been changed at a later stage. Moreover, some of the empirical phenomena these parameters aimed at accounting for have ultimately been shown to depend on other, usually more simple, syntactic properties than the ones they were assumed to be related to. As a consequence, such a subdivision could certainly not have been proposed at the time of their formalization. For all these reasons, however, evaluating the “classical” parameters of the Government-Binding Theory from today’s point of view should allow us not only to verify which kind of parameters has stood the test of time, thus helping us in speculating which aspects of language do allow parameterization and which ones do not, but also to reconsider the notion of parameter and re-evaluate its impact on modern linguistics.

- (1) The main syntactic parameters of the GB phase:
  - (I) Locality parameters:
    - i. *bounding nodes*: Rizzi (1978), Sportiche (1981);
    - ii. *long-distance anaphors*: Manzini & Wexler (1987);
  - (II) Case-assignment parameters:
    - i. *P-stranding*: Kayne (1983);
    - ii. NOM assigned by means of either agreement or government: Koopman & Sportiche (1991);
  - (III) Merge parameters:
    - i. “believe” and *S’ deletion*: Chomsky (1981a);
  - (IV) Linearization parameters:
    - i. *X-bar vs W\* languages*: Hale (1983);
    - ii. *Head-Complement Parameter*: Graffi (1980); Stowell (1981); Travis (1984);
  - (V) Spellout parameters:
    - i. *null subject*: Taraldsen (1978), Rizzi (1982);
    - ii. *V movement to I*: Emonds (1978), Pollock (1989);
    - iii. *V movement to C*: Den Besten (1983);
    - iv. *noun incorporation*: Baker (1988);
    - v. *Overt vs covert wh-movement*: Huang (1982);

### 2.1.1 – Locality, Case-assignment, Merge and Linearization Parameters

Locality parameters concern all those cases in which the possible range of application of a syntactic rule exhibits a certain degree of cross-linguistic variability with respect to the portion of syntactic structure to which the said rule can apply. In the matter in question, the two parameters included in (I) deal with the structural configurations relevant to syntactic movement and binding relations respectively.

Case-assignment parameters deal with cross-linguistic variation in abstract Case assignment. Assuming that syntactic case can be assigned to an NP either *structurally* – that is, simply by virtue of the structural position in which the NP occurs – or *inherently* – namely, depending on the subcategorization properties of the functional head selecting the NP as its complement (cf. Chomsky 1993 [1981a], pp. 170-171); and, further, that «V governs NP in the structural sense [...] but normally P governs NP only in the sense of subcategorization» (Kayne 1983, p. 116), the parameters included in (II) specify whether structural case or inherent case assignment involve proper government by a functional head.

The Merge parametric class corresponds to the homonymous type proposed in Rizzi (2014). As the name suggests, these parameters account for the possible variability in the syntactic category of the complement which can be merged to a functional head of a given type. In this paper, Merge parameters only include the case represented by epistemic verbs of the *believe*-category, which can be merged with infinitive complements labelled either CP(S') or IP(S) and, as a consequence of this binary choice, do or do not allow government of an embedded PRO subject.

Linearization parameters concern the mapping from the syntactic structure projected at a hierarchical level into a linear order for external realization, thus «providing the instructions for the articulatory-perceptual [...] system[s]» (A-P) as required by the nature of the sensory-motor interface (S-M) (Chomsky 1995a, p. 168). According to their original formulation, the parameters included in (IV) can be thought of as being active at different levels: on the one hand, the Head-Complement Parameter specifies the head-complement order at a syntactic level; on the other hand, the Configurability Parameter classifies languages with respect to whether or not their D- and S-Structure levels of representations adhere to the argument-structure configuration originally projected by verbal heads at lexical level. Due to its particularly debated history within the development of Generative Grammar, the Head-Complement Parameter will be reviewed separately in Chapter V.

### 2.1.2 – Some notes on Spellout Parameters

The parametric class proposed under the name of Spellout parameters is partially different from the Spellout class proposed by Rizzi (2014). In addition to those cases in which a base generated empty category can be either licensed or not by a feature

specified on a functional head, as accounted for by the Null Subject Parameter, in the present thesis Spellout parameters also include cases which Rizzi would ascribe to Move parameters, which he defines as expressing the basic syntactic action of Move:

Move parameters express the ability that a head has of attracting another head (incorporation), or a phrase to its specifier position (the latter case being uncontroversial and subsuming the former in some approaches). Parametric properties involving the movement of the verb to an inflectional head [...] and of the inflected verb to the C-system are expressed here, as well as all the parametric variation involved in movement to a Spec position (wh-movement languages vs. wh-*in-situ* languages, etc.). (Rizzi 2014, p. 23)

The different subdivision adopted in this paper, whose possibility was actually briefly hinted at by Rizzi himself, is based on the minimalist distinction between *External Merge* and *Internal Merge* proposed in Chomsky (2001a), which in turn can be regarded as a development of the so-called “Copy Theory” of syntactic movement. While having a long history dating back to the Seventies, as it was notably entertained by Joseph (1976) in order to account for cases of subject-raising in Modern Greek, the Copy Theory of movement has been incorporated into the Minimalist Program since its establishment in Chomsky (1995a) and represents one of the current assumptions of Generative Grammar. According to this approach, an element moved to a higher structural position does not leave in its base position a trace meant as an empty category co-indexed with itself. Rather, the said trace has to be regarded as an actual copy of the moved element, and its absence in overt syntax in case of overt movement is to be imputed not to its vacuous content, but to the fact that the copy’s phonological matrix is deleted in the phonological component before Spellout:

An approach that has occasionally been suggested is the “copy theory” of movement: the trace left behind is a copy of the moved element, deleted by a principle of the PF component in the case of overt movement. But at LF the copy remains [...]. Let us consider this possibility, surely to be preferred if it is tenable. (Chomsky 1995a, p. 202)

Concerning Chomsky’s distinction between *External Merge* and *Internal Merge*, while External Merge basically corresponds to the “canonical” operation Merge by which two distinct syntactic objects are combined in order to form a new syntactic unit, under Internal Merge one of the two syntactic objects which are merged together is a subpart of the other. As a result, not only Internal Merge corresponds *de facto* to syntactic movement



but, crucially, the fact that External Merge and Internal Merge are actually two possible instantiations of the same operation implies that the traditional (transformational) rule “Move  $\alpha$ ” is not different in nature from the most basic operation of syntax: Merge  $\alpha$  and  $\beta$ . As stated by Chomsky:

[Narrow syntax] is based on the free operation Merge. [The Strong Minimalist Thesis] SMT entails that Merge of  $\alpha$ ,  $\beta$  is unconstrained, therefore either *external* or *internal*. Under external Merge,  $\alpha$  and  $\beta$  are separate objects; under internal Merge, one is part of the other, and Merge yields the property of “displacement,” which is ubiquitous in language and must be captured in some manner in any theory. (Chomsky 2004 [2001a], p. 110)

According to this very simple and, therefore, highly desirable account of syntactic movement from a strong minimalist perspective, «Internal Merge leaves a “copy” in place» (*ibidem*). For this reason, the idea that Move represents a subtype of Merge fits well into the approach of Copy Theory, as the application of Internal Merge between a constituent  $\alpha$  and its sub-constituent  $\beta$  does not actually displace  $\beta$ , but rather remerges it into a higher position while leaving a copy of  $\beta$  in its original position.

Returning to the class of Spellout parameters, the approach adopted in the above classification is based on the idea that, assuming both the copy theory of movement and the definition of Internal Merge to hold true, movement is nothing but an epiphenomenon of syntactic operations which, before Spellout, determine the deletion of one of the two copies. In this sense, parameters directly accounting for syntactic movement can be regarded as Spellout parameters as they impose, by means of a strong/generalised EPP feature, the pronunciation of either the higher copy or the lower copy in case of overt and covert movement respectively.

## 2.2 – The formulation of the first explicit parameters

After the “pre-parametric” inquiries conducted by Rizzi and Taraldsen in the late seventies, the systematization of the notion of parameter in Chomsky (1981a) triggered an enormous interest among generative linguists that focused on the proposition that the study of cross-linguistic variation could be crucial in determining the architecture of Universal Grammar (UG) itself. While Generative Grammar’s earlier approach aimed at defining the universal principles of language faculty by relying mainly on the analysis of

English grammar, this new «hope and challenge of simultaneously doing justice to both the similarities and the differences among languages» led to an explosion of comparative research on a wide variety of languages whose momentum has essentially lasted for over three decades to this day (Baker 2011, p. 1).

From this point onwards, the present chapter will be devoted to a thorough review of the main syntactic parameters of the GB Theory by looking primarily at the original works. In this regard, a chronological rather than a classification approach will be followed, particularly for two reasons. First, a chronological review can most likely help us to see how the P&P model has developed over time and how the formulation of syntactic parameters has changed accordingly. Second, as all the individual solutions that had been adopted in accounting for each particular instance of linguistic variation do necessarily reflect the theoretical stage at which they were devised, such an approach would be useful in assessing the role of each parameter within the development of Generative Grammar itself. The next sections (§2.2.1 to §2.2.3) discuss the first three syntactic parameters which characterized the onset of the P&P model, while §2.3 tries to pinpoint the small but important shift in the way parameters are related to principles in Chomsky (1981b). Section 2.4 focuses on some influential proposals which, although often addressed as syntactic parameters within modern syntactic theory, were originally formulated as non-parametric systematic differences. Section 2.5 discusses the formulation of the remaining main parameters of the GB Theory. An overall conclusion follows (§2.6).

### *2.2.1 – The parameterization of Subjacency extended: the bounding nodes in French*

While the growing interest in comparative work led to the discovery of many previously unnoticed systematic cross-linguistic differences, the very first parametric investigations were aimed at extending the descriptive and explanatory scope of those instances of cross-linguistic variation that had been identified between the end of the Extended Standard Theory (EST) phase and the advent of GB-Theory. The first example of this kind of approach is the paper published in 1981 *Bounding nodes in French* by Dominique Sportiche. Starting from the analysis in Rizzi (1978), Sportiche's proposal aimed at accounting for a number of properties which, although displayed systematically in French grammar by the rules of wh-movement and PP-extraposition, not only did behave

differently from those observed in English in the same syntactic contexts, but also could not be explained by means of the set of bounding categories which Rizzi relied on in his account of the locality differences between English and Italian – namely, NP, S and the newly postulated S' (cf. Sportiche 1981, pp. 219-220).

The methodology adopted by Sportiche was simple but effective. Starting from the assumption that all that is necessary for triggering a violation of the Subjacency condition is that a movement rule crosses a *minimum number* of two bounding nodes at a time, his analysis consisted, first, in providing the following list of locality constraints operative in English which, although already formulated in Ross (1967), «are in fact reducible to particular instances of the more general condition [of Subjacency]» (Sportiche 1981, p. 220):

- (2) a. The wh-island constraint
  - b. The upward boundedness of PP-extrapolation (some aspects)
  - c. The complex NP constraint
  - d. The subject condition
- (*ibidem*)

The next step of Sportiche's argumentation was to verify, by means of a set of empirical cases each corresponding to a sentence involving one of the above constraints, whether these four generalizations held true for English were also valid for French. The results of each of these examinations would then be reformulated as a statement providing an explanation for either the grammaticality or the ungrammaticality of the sentence at issue from a bounding perspective.

The first generalization put to test by Sportiche is the "Complex NP Constraint" (CNPC), which is formulated as follows:

Complex NP Constraint: no element contained in a sentence dominated by a noun phrase with a lexical head noun may be moved out of that noun phrase by a transformation.  
(Ross 1986 [1967], p. 76)

In this particular case, the empirical verification seems quite straightforward, as «it is well known that wh-extractions out of a complex NP are impossible in French» as shown by (3) (Sportiche 1981, pp. 221-222):

- (3) \*Qui<sub>i</sub> [<sub>S</sub> connais-tu [<sub>N"</sub> l'homme [<sub>S'</sub> qui a vu t<sub>i</sub> ]]]  
Who do you know the man who saw

Hence, according to Sportiche, the analysis of the CNPC in French in conjunction with the Subjacency principle leads to the postulation of the following statement (Sportiche 1981, p. 222):

- (4) A. At least two out of S', S and N'' are bounding.

The second constraint examined by in this work is the "Subject Condition" (SC), which states «the impossibility of wh-extracting material from inside a subject NP» (*ibidem*):

- (5) \*Which town<sub>i</sub> [<sub>S</sub> do you believe [<sub>S</sub> [<sub>N"</sub> the inhabitants of t<sub>i</sub> ] know this book]]

Although French behaves like English with respect to wh-extractions from inside a subject NP, thus ruling out sentences like (5), in Sportiche's work special attention is paid to the fact that in English the SC is «a special case of a broader generalization [...], namely that no wh-extractions are possible out of any NP» (*ibidem*). While in English this general impossibility is due to the bounding nature of the nodes S and NP triggering a violation of Subjacency, in French movement from inside an object NP does not necessarily cause such a violation, as shown by instances of quantifier extraction as in (6) (Sportiche 1981, p. 223):

- (6) Combien<sub>i</sub> [<sub>S</sub> as tu vu [<sub>N"</sub> t<sub>i</sub> de personnes]]  
How many did you see (of) persons

In French, what does indeed cause ungrammaticality is to perform quantifier extraction out of PP's, as in (7) (Sportiche 1981, p. 224):

- (7) \*Combien<sub>i</sub> [<sub>S</sub> as tu voté [<sub>PP</sub> pour [<sub>N"</sub> t<sub>i</sub> de démocrates]]]  
How many did you vote for (of) democrats

Considering both examples (7) and (8) from a locality perspective, according to Sportiche their different grammatical status leads to the formulation of statements (B1) and (B2) respectively (Sportiche 1981, pp. 223-224):

- (8) B1. At most one out of S and N'' is bounding.  
B2. At least two out of PP, S and N'' are bounding.

In the wake of the above considerations regarding the possibility of moving a constituent out of a complement, in his analysis Sportiche took into account also some different cases of noun complement extraction. On the one hand, in French there is no ban on extraction of NP complements of nouns, as in example (9) «the wh-phrase appearing in COMP position [...] has been moved from the position that noun complements usually occupy in deep structure» (Sportiche 1981, pp. 224-225):

- (9) Le livre dont<sub>i</sub> [<sub>S</sub> il connaît [<sub>N''</sub> la fin t<sub>i</sub> ]]  
The book of which he knows the end...

On the other hand, however, the same operation applied to PP complements is not always possible. Leaving aside PP complements introduced by prepositions other than genitive *de*, which in general are not wh-extractable, Sportiche focused his attention on genitive PP's, which display a non-uniform behaviour depending on their syntactic context (cf. Sportiche 1981, p. 225). For example, while a genitive PP is wh-extractable from an NP if the latter is immediately dominated by a node S, «extraction of a complement is impossible if the larger NP is itself embedded inside a PP» (Sportiche 1981, p. 227). The outcomes of these two operations are shown by (10) and (11) respectively (*ibidem*):

- (10) Ce théorème, dont<sub>i</sub> [<sub>S</sub> [<sub>N''</sub> une démonstration t<sub>i</sub> ] les a convaincu de l'utilité de ces techniques d'analyse]]...  
This theorem, of which a proof convinced them of the use of these techniques of analysis...
- (11) \*Ce théorème, dont<sub>i</sub> [<sub>S</sub> il a travaillé [<sub>PP</sub> sur [<sub>N''</sub> une [<sub>N'</sub> démonstration t<sub>i</sub> ]]]]]...  
This theorem, of which he worked on a proof...

Similarly, considering the genitive clitic form *en*, «the application of a movement rule moving the clitic *en* from the adnominal complement position it occupies in D-structure to the clitic position» is possible only as long as the adnominal complement position it occupies in D-structure is not embedded inside a PP (Sportiche 1981, p. 228). Accordingly, sentence (12) allows clitic movement, while sentence (13) does not (Sportiche 1981, pp. 227-228):

- (12) Il  $en_i$  a lu  $[_{N''} \text{ la } [_{N'} \text{ demonstration } t_i ]]$   
 He of-it read the proof
- (13) \*Il  $en_i$  a travaillé  $[_{PP} \text{ sur } [_{N''} \text{ la } [_{N'} \text{ demonstration } t_i ]]]$   
 He of-it worked on the proof

Putting together all the data obtained from examples (10), (11), (12), and (13), according to Sportiche the following propositions must hold (Sportiche 1981, p. 228):

- (14) C1. At most one out of  $N''$  and  $N'$  is bounding (since [10] is grammatical).  
 C2. At least two out of PP, S,  $N''$  and  $N'$  are bounding (because of the ungrammaticality of [11]).

The third means by which Sportiche aimed at determining the set of bounding categories in French is the process of PP-extrapolation (PP-ex). In order to make sure to be concerned exclusively with cases of PP-ex which are truly susceptible to Subjacency (and which are not, for example, simply stylistic) only sentences as the following ones are taken into account, in that they involve a relation between an extraposed PP and «an empty category necessary to act as an “antecedent” (i.e., a source under a movement hypothesis) to the extraposed phrase» (Sportiche 1981, pp. 228-229):

- (15) Un commentaire de ce livre vient de paraître.  
 A commentary of this book just appeared
- (16)  $[_{N''} \text{ Un } [_{N'} \text{ commentaire } t_i ]]$  vient de paraître  $[_{PP} \text{ de ce livre}]$

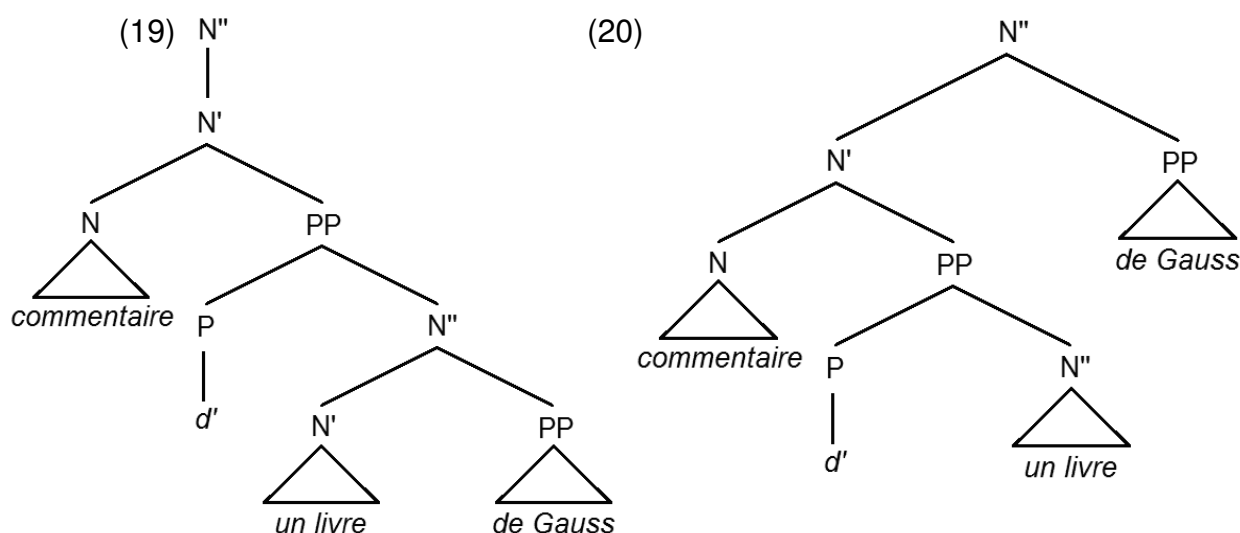
Since example (16) constitutes a well-formed sentence, according to Sportiche the possibility of extraposing a PP out of a simple NP leads to the formulation of statement D1 (Sportiche 1981, p. 230):

- (17) D1. At most one out of N' and N'' is bounding.

In order to provide a syntactic context in which it is not possible to perform the same operation due to a violation of Subjacency, Sportiche resorts to the following paradigm (Sportiche 1981, p. 229):

- (18) a. Un commentaire d'un livre de Gauss vient de paraître.  
 A commentary of a book by Gauss just appeared  
 b. Un commentaire d'un livre vient de paraître de Gauss.

While on the one hand in (18a) an ambiguity arises between the two plausible readings “Gauss wrote a commentary of a book” and “Gauss wrote a book”, on the other hand in (18b) only the former is available. These two interpretations, in turn, can be structurally represented as in (19) and (20) respectively (Sportiche 1981, p. 230):



According to Sportiche, structure (19) does not allow PP-ex of the most embedded PP, the reason being that there are too many bounding nodes intervening: PP, N' and N'', the latter of which occurs twice. At this point, there are two distinct statements which would account for the impossibility of PP-ex in (19) (*ibidem*):

- (20) D2. At least PP and N' are bounding; or  
 D3. At least N'' is bounding.

The last generalization put to test in Sportiche's paper is the "Wh-island constraint", which is the generalization by means of which Rizzi (1978) had suggested that the set of bounding nodes in Italian is different from that of English. In his comparison between English and French, Sportiche observed that, «contrary to what happens to English, [...] in French, like in Italian, violations of wh-islands are grammatical under certain circumstances», as shown by (21) (Sportiche 1981, pp. 232-233):

- (21) C'est à mon cousin [<sub>S'2</sub> que<sub>i</sub> [<sub>S2</sub> je sais [<sub>S'1</sub> lequel<sub>j</sub> [<sub>S1</sub> offrir t<sub>j</sub> t<sub>i</sub> ]]]]  
 It is to my cousin that I know which one to offer

By means of this example, Sportiche essentially reposed for French the same argument put forth by Rizzi for Italian. More precisely, the only way for (21) to not violate neither the strict cyclicity condition nor the doubly-filled COMP filter is to assume that «S' rather than S is bounding» (Sportiche 1981, pp. 233, 235):

- (22) E. S' is bounding and S is not.

The results of Sportiche's systematic empirical inquiry regarding the validity of the constraints *a*, *b*, *c* and *d* in French is represented by the statements A to E that have been listed so far. For the sake of simplicity, only the most relevant ones are repeated below:

- (23) A. At least two out of S', S and N'' are bounding.  
 B2. At least two out of PP, S and N'' are bounding.  
 E. S' is bounding and S is not.

The point of this list is that, because of E, the minimal set of bounding nodes compatible with both A and B2 (and all the other statements) must be {S', N'', PP}. According to Sportiche, this conclusion confirms that, as envisioned in Rizzi (1978), «parametrization of the set of bounding categories appears to be necessary» in order to uphold unchanged the status of Subjacency as a universal principle of UG (Sportiche 1981, p. 220).



### 2.2.2 – The formulation of the Null Subject Parameter

Although an important correlation between null subjecthood and other typological features had been already noted in Taraldsen (1978), which postulated a link between the occurrence of silent definite pronominal subjects, rich agreement and the possibility of extracting a subject through an overt complementizer, the Null Subject Parameter was first formulated by Rizzi (1982). In his paper *Negation, Wh-movement and the null subject parameter*, Rizzi not only did subsume the main ideas which had been proposed in some recent studies in the area of null subject languages, but also reinterpreted their results in order to derive, from an individual parameter setting, the following clustering of syntactic properties (Rizzi 1982, p. 117):

- (24) a. phonetically null subjects in tensed clauses
- b. free process of subject inversion
- c. apparent violation of the COMP-t effect

Starting from this set of linguistic features, the typological opposition between null subject and non-null subject languages is in turn exemplified by Italian and English, which display the mutually disjoint sets of options (a) and (b) (*ibidem*):

- (25) a. *e* verrà
- b. \**e* will come
- (26) a. *e* verrà Gianni
- b. \**e* will come Gianni
- (27) a. Chi<sub>i</sub> credi che *e*<sub>i</sub> verrà?
- b. \*Who<sub>i</sub> do you think that *e*<sub>i</sub> will come?

Rizzi's argumentation begins by declaring his own intent to reformulate Taraldsen's proposal, together with other ideas about null subject languages that would later be reinterpreted in this paper, in the light of the "Empty Category Principle" (ECP), a universal constraint on the possible occurrence of empty categories which was originally formulated in Chomsky (1981a):

- (28) The Empty Category Principle: [<sub>α</sub> e] must be properly governed.  
(Chomsky 1993 [1981a], p. 250)

Unlike the ECP, the Nominative Island Condition (NIC) commonly referred to as the cause of ungrammaticality in English tensed sentences featuring non-overt subjects was actually a spurious generalization which could not account for «the “English side” of the parameter», as put forth by Rizzi, in a consistent way (Rizzi 1982, p. 118). According to Chomsky (1981a), while the NIC could potentially account for the ungrammaticality of unbounded subject anaphors in sentences like (25b) and (26b), traces of wh-movement like that in (27b), not being anaphors but variables, should not be subject to the NIC, as they should be free from any antecedent (cf. Chomsky 1993 [1981a], p. 231). Accordingly, «the conclusion that variables are not subject to the NIC» suggests that «[the phenomenon ruling out wh-traces in subject position], while similar to the NIC effects, is really a different phenomenon» – namely, the ECP (Chomsky 1993 [1981a], p. 232):

Chomsky (1981a) has convincingly shown that the NIC, conceived of as a binding principle for anaphors, was a spurious generalization, and that its effects should be factored out differently. In the system of Chomsky (1981a), the general principle which subsumes the empirical coverage of NIC as far as the “null subject” effects are concerned is the Empty Category Principle (ECP) [...]. (Rizzi 1982, p. 118)

In these terms, as «the subject position of a tensed sentence in English is not a properly governed position», the ECP systematically rules out not only sentences like (25b), but «the same account holds when the subject position is transformationally vacated via rightward NP movement [...] or via wh-movement of the subject across an overt complementizer», as in (26b) and (27b) respectively (Rizzi 1982, pp. 118-119).

In order to prepare the ground for the formulation of the null subject parameter, Rizzi presents an important observation which had been previously noted by Kayne (1981). In French, the negative indefinite *personne* requires the presence of the negative marker *ne*. According to Kayne’s data, «if *personne* is in object position of an embedded clause, *ne* can be cliticized onto the embedded verb (as in [29a]) or, somewhat more marginally, onto the main verb (as in [29b])» (Rizzi 1982, p. 119):

- (29) a. J’ai exigé qu’ils *n’arrêtent personne*.  
“I have required that they *neg* arrest *nobody*.”

- b. ?Je n'ai exigé qu'ils arrêtent *personne*.  
 "I *neg* have required that they arrest *nobody*."

However, while «there is [...] no general prohibition against negative "personne" in embedded subject position» (Kayne 1983 [1981], p. 24), in this particular case the *ne* cannot occur in the main clause, but only in the embedded one (*ibidem*):

- (30) a. J'ai exigé que *personne ne* soit arrêté.  
 "I have required that *nobody neg* be arrested."  
 b. \*Je n'ai exigé que *personne* soit arrêté.  
 "I *neg* have required that *nobody* be arrested."

Kayne's intuition was that «in the "ne"..."personne" construction, "ne" reflects the scope of "personne" which we take to be quantifier-like» (*ibidem*). Therefore, as shown by (29) and (30), either narrow scope (as in 29a) or wide scope (as in 29b) interpretation is possible for the quantifier-like negative element *personne* in the post-verbal position of the embedded clause. On the contrary, as shown in (30b), the possibility of wide scope with *personne* in embedded subject position is impossible. In accordance with May's (1977) format, scope is assigned to quantifier phrases in Logical Form (LF) by "Quantifier Raising" (QR), which consists in a movement rule adjoining the quantified phrase to the left of the S over which the phrase in question is interpreted as having scope. Accordingly, the respective LF's of (29) and (30) would be (31) and (32) (Rizzi 1982, p. 120):

- (31) a. [<sub>S'</sub> [<sub>S</sub> J'ai exigé [<sub>S'</sub> que *personne*<sub>i</sub> [<sub>S</sub> ils arrêtent *e*<sub>i</sub> ]]]]  
 b. [<sub>S'</sub> *personne*<sub>i</sub> [<sub>S</sub> j'ai exigé [<sub>S'</sub> qui [<sub>S</sub> ils arrêtent *e*<sub>i</sub> ]]]]  
 (32) a. [<sub>S'</sub> [<sub>S</sub> J'ai exigé [<sub>S'</sub> que *personne*<sub>i</sub> [<sub>S</sub> *e*<sub>i</sub> soit arrêté ]]]]  
 b. \*[<sub>S'</sub> *personne*<sub>i</sub> [<sub>S</sub> j'ai exigé [<sub>S'</sub> que [<sub>S</sub> *e*<sub>i</sub> soit arrêté ]]]]

The point of the asymmetry between (31) and (32) is that, in French, sentences like (30b) are ungrammatical because the trace of *personne* in embedded subject position is not properly governed at LF, as shown in (32b). On the contrary, in both LF's (31a) and (32b) the trace is properly governed by either the quantifier itself or the embedded verb, hence both (29a) and (29b) are well formed.

At this point, it would seem plausible to assume in advance that, all other things being equal, in null subject languages like Italian sentences corresponding to (30b) should be rescued from ungrammaticality because, as stated by Taraldsen (1978), these languages' rich agreement inflection would act as a proper governor. Therefore, the next step of Rizzi's argumentation was to verify if the asymmetry noted by Kayne in French would also hold for Italian. Contrary to the previous assumption, however, in Italian the embedded quantified subjects cannot take wide scope, exactly as in French. Rizzi came to this conclusion by examining the following sentences, which correspond to (30a) and (30b) respectively (Rizzi 1982, p. 124):

- (33) a. *Non pretendo che tu arresti nessuno.*  
 “(I) *neg* require that you arrest *nobody*.”  
 b. *Non pretendo che nessuno ti arresti.*  
 “(I) *neg* require that *nobody* arrest you.”

In sentence (33a) the wide scope interpretation of *nessuno* is acceptable, as in its FL the empty NP resulting from QR is properly governed by the embedded verb (*ibidem*):

- (34) [<sub>S'</sub> *non* + *nessuno*<sub>i</sub> [<sub>S</sub> pretendo [<sub>S'</sub> che tu arresti *e*<sub>i</sub> ]]]  
 “There is no *x* such that I require that you arrest *x*”

Being well-formed, (33b) seems to deny the existence, in Italian, of the wide scope asymmetry that Kayne noticed in French. However, if this were actually the case, in this sentence the quantifier-like element *nessuno* should be able to take wide scope, but it is not. According to Rizzi «[33b] is only acceptable in the totally irrelevant interpretation explicitly represented in LF [35]», that is, the one with «*nessuno* receiving narrow scope interpretation» (Rizzi 1982, p. 125):

- (35) [<sub>S'</sub> *non* pretendo [<sub>S'</sub> che *nessuno*<sub>i</sub> [<sub>S</sub> *e*<sub>i</sub> ti arresti]]]  
 “I do not require that there be no *x* such that *x* arrest you”

In fact, for (33b) the relevant interpretation with *nessuno* receiving wide scope interpretation is impossible (*ibidem*):

- (36) \* $[_S \text{ non} + \text{nessuno}]_i [_S \text{ pretendo } [_S \text{ che } [_S e_i \text{ ti arresti}]]]$   
 “There is no x such that I require that x arrest you”

Despite that fact that Italian is a null subject language, example (36) shows that neither in non-pro drop languages nor in pro-drop ones INFL can license a trace in embedded subject position. The subsequent need to revise the actual role of agreement inflection in the Null Subject Parameter differently led Rizzi to reconsider what «seems to be a good argument, internal to the Italian syntax, in favor of the relevance of INFL for the well-formedness of “null subject” sentences» (Rizzi 1982, p. 127). The first step in this direction was to single out some Italian sentences in which a tensed inflection with overt morphological agreement does not by itself determine the well-formedness of a zero subject (Rizzi 1982, p. 128):

- (37) a. [Essendo piovuto per tutto il pomeriggio,] non siamo usciti.  
 “[Having rained for the whole afternoon,] we didn’t go out.”  
 b. Ritengo [esser nevicato anche sotto i mille metri]  
 “I believe [to have snowed even below 100 meters]”

There are two interesting observations that can be made from the above examples. First, in the non-finite gerund and infinitival clauses highlighted in (37) «a null subject is never possible with the interpretation of a definite pronoun, but is possible [...] with the “dummy” interpretation corresponding to French *il*, English *it* and/or *there*» (Rizzi 1982, pp. 128-129). According to Rizzi, the acceptability of null subjects in such uninflected clauses is likely to be related to the positive setting of the Null Subject Parameter, as sentences corresponding to (37) in non-pro drop languages such as English and French are ungrammatical. Second, in these uninflected clauses the auxiliary is moved to the sentence-initial position, which it is assumed to be COMP position. This construction is not unique to the occurrence of “dummy” null subjects, as it affects also uninflected clauses with a definite pronoun in subject position (Rizzi 1982, p. 129):

- (38) a. Avendo tu telefonato alla polizia,...  
 “Having you telephoned to the police,...”  
 b. Ritengono [non esser io in grado di affrontare la situazione]  
 “They believe [not-to-be I able to face the situation]”

(38a) and (38b) are useful to understand an important aspect of AUX movement to COMP, as their grammaticality implies that this property does compensate for the absence of tensed inflection, which in ordinary tensed clauses assigns nominative case to the subject under government. More precisely, in (38a) and (38b) there is no violation of the “Case Filter”, a universal principle requiring lexical NPs to be case marked which was originally formulated in Chomsky (1980):

- (39) The Case Filter: \*N, where N has no Case  
(Chomsky 1980, p. 25)

Putting together these two observations, Rizzi postulated the following generalizations: on the one hand, contrary to what was commonly assumed, the role of rich agreement inflection is not crucial in licensing phonetically null subjects, but simply determines the possibility of a definite/referential interpretation; on the other hand, although referential null subjects are not possible in uninflected clauses, expletive/non-referential null subjects are licensed if associated with a nominative case position (Rizzi 1982, p. 130):

- (40) (I) A phonetically null subject with “dummy” interpretation can be found in the local context of a nominative assigner (tensed inflection or Aux in COMP).  
(II) A phonetically null subject with definite pronominal interpretation can be found in the local context of a tensed inflection.

Another important aspect which sets apart Rizzi (1982) from other previous accounts of the difference between pro-drop and non-pro drop languages concerns the way this discrete typological variation is conceptualized in terms of parametric features. Rizzi’s proposal consisted in a modification of the so-called “Chomsky-Taraldsen solution”, whose aim was to provide an elegant explanation about the parallel syntactic relations occurring between both subject NP and verbal inflection and object NP and object clitic (*ibidem*):

- (41)  $[_{NP} e_i] t_j \text{ conosco } [_{NP} e_j]$   
“ e you + know e ”  
“I know you”

Since both verb inflection and object clitic behave in very much the same manner in being able to bind and properly govern their respective empty argument positions, in null subject languages the verbal affix would act exactly as an object clitic in its ability to license a null subject. However, from this point onwards, Rizzi's account diverges from this solution with respect to the role attributed to rich inflection. On the one hand, while in Taraldsen (1978) it was assumed that the special device capable of licensing an empty NP position in null subject languages was morphologically rich inflection itself, on the other hand Rizzi hypothesised that the well-formedness of structures like (41) were due to the pronominal status of inflection. In this sense, according to Rizzi, phonetically null subjects could only be bound and properly governed by a functional head INFL being specified with the feature [+ pronoun] (cf. Rizzi 1982, p. 130):

[...] The most direct way of capturing the parallelism in [41] would consist in suggesting that the characteristic property of NSL's is that their verbal inflections have (clitic-like) pronominal properties. This intuition can be straightforwardly implemented by assuming that INFL in NSL's is specified with the feature [+ pronoun]: i.e., like a clitic, it is a verbal affix with (pro-) nominal properties, specified with respect to such grammatical features as person and number; and, like a clitic, it is interpreted as a definite pronoun (at least in structures like [41]), and binds and properly governs an empty NP position. (Rizzi 1982, pp. 130-131)

In addition to the important theoretical advancement resulting from having formalized a maximally simple and discrete format for the (parametric) difference distinguishing pro-drop languages from non pro-drop ones, the characterization of the Null Subject Parameter as a binary feature [ $\pm$  pronoun] specified on INFL had also the advantage of accounting for an important difference between subject and NP inflection and object NP and object clitic. In fact, despite the parallelism shown in (41), «the verbal inflection can obviously co-occur with a lexical subject, while an object clitic cannot co-occur with a lexical object (putting aside cases of so-called “clitic doubling”)» (Rizzi 1982, p. 131):

- (42) (Mario<sub>i</sub>) lo<sub>j</sub> conosce<sub>i</sub> (\*Gianni<sub>j</sub>)  
 “(Mario<sub>i</sub>) him<sub>j</sub> + knows<sub>i</sub> (\*Gianni<sub>j</sub>)”

As Rizzi points out, «the lack of co-occurrence between object clitics and object NP's is accounted for by the Case component» (*ibidem*). Starting with the assumption that «a Case assigner can assign Case only once», in a structure like (42) the Case Filter

would always be violated because either the clitic or the object NP would be assigned accusative Case by the verb, but not both (Rizzi 1982, p. 131). However, such a scenario is never found with lexical subjects and INFL precisely due to the nature of the feature [+pronoun] which, according to Rizzi, while not possible in non-null subject languages it is optional in null subject languages:

Why is such a complementarity never found with subject NP's and INFL? I would like to suggest that this is to be attributed to the fact that the inflection of NSL's is optionally specified [+pronoun]. If so specified, INFL has clitic-like properties, pronominal interpretation, and must absorb nominative Case because of the Case Filter. If it is not, it does not differ in any significant respect from the verbal inflection of non-NSL's. (*ibidem*)

The immediate consequence of the fact that INFL is not always specified [+pronoun] in null subject languages is that, crucially, when verbal inflection co-occurs with a lexical subject, INFL is always specified [-pronoun]. In this way, nominative Case has to be assigned only to the lexical subject, and therefore there is no violation of the Case Filter.

Besides providing an account of the impossibility of structures like (42), the definition of the Null Subject Parameter in (40) also plays a crucial role in accounting for the fact that, while wide scope interpretation is available for quantifier-like negative elements post-verbal position of the embedded clause, embedded quantified subjects cannot take wide scope in both French and Italian. As has already been said, in sentences like (33b) (here repeated as (43)), *nessuno* should move outside S' in Logical Form to receive wide scope, but by doing so its own trace is left ungoverned and the resulting configuration triggers a violation of the ECP:

- (43) \* $[_{S'} \text{non} + \text{nessuno}_i [_S \text{pretendo} [_{S'} \text{che} [_S e_i \text{ti arresti}]]]$   
 "There is no x such that I require that x arrest you"

According to the proposed treatment of the Null Subject Parameter, the only means to save the derivation from ungrammaticality would be to have INFL specified with the feature [+pronoun] to properly govern the empty NP. However, although on the one hand this specific operation would in fact satisfy the ECP, on the other hand «INFL [...] would absorb Case, and no Case would be assigned to the lexical subject, in violation of the Case Filter» (Rizzi 1982, p. 138). In conclusion, if in French the impossibility of having



INFL specified with [+ pronoun] directly excludes the application of QR to a pre-verbal, embedded subject, in Italian the pronominal nature of verbal inflection potentially satisfies the ECP, but nonetheless checkmates the Case component in a similar way as in (42) (except from the fact that, in (43), it is nominative case that cannot be properly assigned, not accusative case).

From what has been argued so far, Rizzi was finally able to express the Null Subject Parameter by means of the two following properties:

- (44) The Null Subject Parameter:
- a. INFL can be specified [+ pronoun].
  - b. INFL [+ pronoun] can be referential.
- (Rizzi 1982, p. 143)

Within Rizzi's treatment, (44a) and (44b) are to be considered as two related but autonomous properties, as the potential occurrence of the latter strictly depends on the positive value of the former in such a way that « property [44b] can only be found in a subset of the cases in which property [44a] holds» (*ibidem*). On the one hand, (44a) accounts for the fact that a phonetically null subject can only be licensed by a pronominal INFL – that is, the feature-driven parametric property which crucially distinguishes pro-drop languages like Italian from non-pro-drop languages like English and French. On the other hand, (44b) accounts for the descriptive statement (40II) as it «is directly related to the actual morphological richness of the inflection, while property [44a] is not» (*ibidem*). As seen previously with the uninflected clauses in (37), the morphological richness of INFL – which Rizzi assumes to «be referential only if specified [ $\alpha$  person] ( $\alpha$  ranging over I, II, III)» – is not directly able to license a null subject, but only adds to its interpretation by identifying it as referential or not (Rizzi 1982, p. 142).

In the second part of the paper, Rizzi aimed at utilizing the newly proposed definition of the Null Subject Parameter in (44) to account for the typological properties (24b) and (24c), which were identified as co-occurring with the possibility of null subjects in tensed clauses (24a). The first property to be inquired was free subject inversion in simple sentences, which in turn was assumed to be an instance of subject adjunction to the VP node (Rizzi 1982, p. 132):

- (45)  $e_i$  INFL<sub>i</sub> [<sub>VP</sub> [<sub>VP</sub> ha telefonato] Gianni<sub>i</sub>]  
 “ $e_i$  INFL<sub>i</sub> has telephoned Gianni<sub>i</sub>”

Leaving aside further consideration of the Case relation between the subject NP position and the postverbal lexical subject, there are two main aspects in the derivation of sentence (45) which need some explanation. More precisely, the first question that arises is how this sentence can be well-formed with respect to the ECP; secondly, the other question is how the same sentence can be well-formed with respect to the binding principle (cf. Rizzi 1982, p. 132).

Regarding the first question, according to Rizzi the answer is pretty simple, as the well-formedness of (45) follows directly from the status of INFL in Italian: INFL, being optionally [+ pronoun] in null subject languages, behaves like a pronoun in all relevant respects and can thus properly govern the subject's trace.

As far the second question is concerned, the other problem pointed out by Rizzi is the fact that a sentence like (45) seems to contradict both Principle B and Principle C of the Binding Theory laid out in Chomsky (1981a):

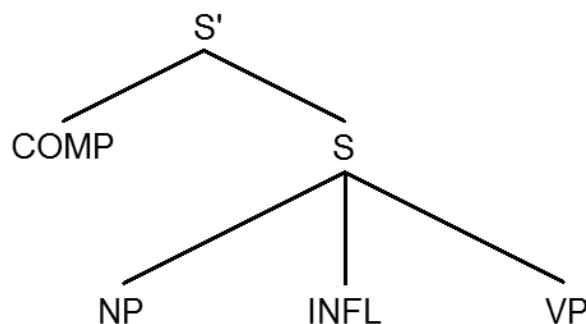
(46) Binding Theory:

- A. An anaphor is bound in its governing category.
- B. A pronominal is free in its governing category.
- C. An R-expression is free.

(Chomsky 1993 [1981a], p. 188)

First, the subject trace seems to bind the pronominal INFL, thus violating Principle B. Although seemingly strange in light of the current theoretical assumptions of Generative Grammar, this consideration was actually plausible for Rizzi since, before the *Barriers* framework formulated in Chomsky (1986a), the subject's base position and INFL were assumed to be sister nodes, thus symmetrically c-commanding each other:

(47)



Second, also Principle C seems to be at stake, and this because the post-verbal subject *Gianni*, a referential expression, appears to be bound by both the empty subject and the pronominal INFL.

In order to justify (43) with reference to the binding principle, Rizzi's focused his attention on «how the binding relation interacts with the thematic structure of a sentence» to find a specific syntactic configuration which could be assumed to exclude any binding relation between two thematic positions (Rizzi 1982, p. 135). The specific cases analyzed by Rizzi were the following (*ibidem*):

- (48) (a) My friends<sub>i</sub> helped each other<sub>i</sub>.  
(b) My friends<sub>i</sub> were helped e<sub>i</sub>.

In (48a) the binding element *my friends* and the bound element *each other* have independent thematic roles, as the first is an Agent and the second is a Theme. On the other hand, in (48b) the binding element depends on the bound element for its own thematic role, as *my friends* receives the theta-role Theme from the trace in object position. While not directly related to (43), Rizzi notes that the two examples (48a) and (48b) actually represent all the available analyses of theta structure ever discussed in the literature. More precisely, while (48a) and (48b) do not cover alone all theoretically possible interactions between binding relations and thematic structure, as there still is «a third a priori possible case [...] i.e., the case in which the binding element transmits its  $\theta$  role to the bound element», this latter configuration is practically unattested and, according to Rizzi, this lack of reference to it does directly argue for its impossibility (Rizzi 1982, pp. 135-136). Crucially, Rizzi observed that this never-discussed, not-even-verbalized configuration does indeed correspond to the problematic cases triggering violations of Principles B and C in free subject inversion:

Now, if we carefully consider all the problematic cases discussed so far, we notice that they have this form exactly: there is an element  $\alpha$  which c-commands and is coindexed with an element  $\beta$ , and  $\beta$  receives its  $\theta$  role from  $\alpha$ . In exactly these cases, the theory of binding would require that  $\alpha$  not count as a binder of  $\beta$ , even in spite of the fact that the structural requirement of the definition of the binding relation is met. (Rizzi 1982, p. 136)

Therefore, Rizzi's answer to the previous objections raised against the well-formedness of (45) was to conclude that «the notions “ $\alpha$  binds  $\beta$ ” and “ $\beta$  is  $\theta$  dependent

from  $\alpha$ ” (i.e.,  $\beta$  receives its  $\theta$  role from  $\alpha$ ) seems to be mutually incompatible» (Rizzi 1982, p. 136), as captured by the following reformulation to the definition of the binding relation:

- (49) “ $\alpha$  binds  $\beta$  iff  $\alpha$  c-commands and is coindexed with  $\beta$ , and  $\beta$  is not  $\theta$  dependent from  $\alpha$ .” (*ibidem*)

This revision accounts for all the otherwise ungrammatical configurations concerning subject inversion’s example (43), repeated below as (50):

- (50)  $e_i$  INFL <sub>$i$</sub>  [<sub>VP</sub> [<sub>VP</sub> ha telefonato] Gianni]  
 “ $e_i$  INFL <sub>$i$</sub>  has telephoned Gianni <sub>$i$</sub> ”

According to (49), in (50) the pronominal INFL cannot be bound by the empty subject, as the former is dependent from the latter as far as its thematic role is concerned. Similarly, the postverbal subject cannot be bound by either the pronominal INFL or the empty subject. While on the one hand *Gianni* directly depends on INFL for its own thematic role and thus cannot be bound by it, on the other hand the postverbal subject is theta-dependent also from the empty subject, although in an indirect way, if we assume that «the relation “ $\beta$  is  $\theta$  dependent from  $\alpha$ ” is transitive» (*ibidem*). This latter analysis is in turn based on the assumption that, as argued by Rizzi, the Agent theta-role is transmitted from tensed INFL to the VP-adjoined subject, together with nominative Case, by means of the procedure illustrated in (51), which is feasible on condition that pronominal INFL acts as an expletive/non-referential subject licenser:

- (51) in the structure  
 ...dummy <sub>$i$</sub> ...NP <sub>$i$</sub> ...  
 where NP <sub>$i$</sub>  is coindexed with and in the domain of dummy <sub>$i$</sub> , copy the Case of  
 dummy <sub>$i$</sub>  on NP <sub>$i$</sub> .  
 (Rizzi 1982, p. 133)

Rizzi’s account of free subject inversion played a fundamental role in the treatment of the Null Subject Parameter. In fact, while seeming to be a mere surface effect, this property of null subject languages was also argued to be one of the structural steps needed by a grammar in order to perform subject wh-extraction. This consideration makes

us turn to the last property in the typological cluster of null subject languages, that is, the apparent violation of the COMP-t effect with *wh*-movement:

(52) Chi<sub>i</sub> credi che e<sub>i</sub> verrà?

\*Who<sub>i</sub> do you think that e<sub>i</sub> will come?

According to Rizzi, *wh*-sentences like (52) are particularly problematic as they potentially raise two problems. First, from a descriptive perspective, the fact that (52) is well-formed in Italian but not in a language like English or French could in fact suggest that, while in both languages quantifier movement could be assumed to behave in a similar way with respect to COMP-trace effect, the same does not seem to be true also for *wh*-movement. As pointed out by Rizzi:

[...] such examples as [52] seem to indicate that in Italian *wh* constructions do not show COMP-trace effects: we would therefore be led to the conclusion that quantified constructions in French and Italian, and *wh* constructions in French are to be incorporated within the same descriptive generalization (the COMP-trace effect), while *wh* constructions in Italian pattern differently (Rizzi 1982, p. 146)

Second, from a more technical point of view, at first sight the well-formedness of (52) seems unaccountable by means of the theoretical model assumed so far. Consider what Rizzi describes as «the most straightforward derivation of [52]» (*ibidem*):

(53) [<sub>COMP</sub> Chi<sub>i</sub>] credi [<sub>S'</sub> che e<sub>i</sub> INFL<sub>i</sub> verrà]

Similarly to (36), (53) is in fact ill-formed regardless of the value of [ $\pm$  pronoun] on INFL. On the one hand, [- pronoun] INFL would be incapable of properly governing the trace left by *wh*-movement, thus violating the ECP. On the other hand, [+ pronoun] INFL would not only properly govern the trace, but also bind it. Accordingly, since traces of *wh*-movement are variables, the empty NP position in (53) should be antecedent-free and not coindexed with a pronominal in subject position. In other words, from (53) it invariably follows that subject extraction should be impossible also in null subject languages:

There are two cases, according to whether the embedded INFL is pronominal or not. If it is non-pronominal, the structure is ruled out by ECP, as its English counterpart. If INFL is pronominal, then ECP is not violated, since the trace is properly governed; but INFL also

binds the empty subject position which, being a *wh* variable, should be free because of clause (C) of the binding principle. Hence, the structure is ruled out by the theory of binding. (Rizzi 1982, p. 146)

However, according to Rizzi it is exactly the asymmetrical behaviour between preverbal and postverbal subjects position with reference to wide scope interpretation which suggests that (53) does not represent the correct derivation of (52). More precisely, the solution adopted by Rizzi was to assume that the position from which the embedded subject is extracted by *wh*-movement is not its base position, but the post-verbal one:

[...] taking seriously the parallelism between scope assignment and *wh* movement, since post-verbal subjects only can be extracted from clauses by the scope assignment procedure, it is reasonable to consider the possibility that, in a parallel fashion, postverbal subjects only can be *wh* extracted. (Rizzi 1982, pp. 146-147)

Following this reasoning, the correct derivation of (52) is represented in (54) (Rizzi 1982, p. 147):

(54) [COMP Chi<sub>i</sub>] credi [<sub>S</sub>' che e<sub>i</sub> INFL<sub>i</sub> verrà e<sub>i</sub>]

This solution implies that, in null subject languages, *wh*-extraction crucially takes place after subject inversion, as it is movement of the embedded subject to post-verbal position which ensures that the two problems pointed out above for (52) do not occur. On the one hand, there is no need to postulate any descriptive asymmetry between quantified constructions and *wh*-constructions in French and Italian. In fact, it follows from (54) that QR and *wh*-extraction are impossible from subject position in both French and Italian. On the other hand, by assuming that in sentences like (52) INFL is specified [+ pronoun], the derivation of (54) is immune from any technical problem. Since INFL is pronominal, the subject trace in pre-verbal position is properly governed. Moreover, while being properly governed by the verb, according to (51) the trace in object position (a variable) is also free, as it is transitively theta-dependent from both INFL and the preverbal trace (cf. *ibidem*).

In conclusion, Rizzi (1982) not only represents an innovation with respect to the previous assumptions about the Null Subject Parameter, but it is also an important contribution to parametric theory in its globality. In addition to the idea that the “null subject” phenomenology directly embraces both the licensing mechanism and the possible (±referential) interpretation of null subject pronouns, the system of features adopted in

describing these particular instances of linguistic variation would subsequently help Generative Grammar in outlining a discrete and univocal format to parametric variation.

### 2.2.3 – *The lexical parameter of S'-deletion*

One, if not the main, goal pursued in Chomsky's seminal work *Lectures on Government and Binding* (1981a) was the development of Binding Theory and, as a corollary, the formulation of a typology of NP categories both with and without phonological content. Overt NPs were distinguished into the three types which are expressly referred to in the three universal principles of Binding Theory, that is, anaphors, pronominals and referential expressions:

(55) Binding Theory:

- D. An anaphor is bound in its governing category.
- E. A pronominal is free in its governing category.
- F. An R-expression is free.

(Chomsky 1993 [1981a], p. 188)

As far as covert NP categories are concerned, according to Chomsky (1982) the empty counterparts of (lexical) anaphors, pronominals and nouns were, in order, NP-traces, *pro*, and *wh*-traces (the latter being referred to as variables), as they share their binding behaviour with their respective overt equivalent. However, differently from overt NPs, still one type of covert NP category remains: *PRO*, that is, the base-generated subject position of infinitive clauses, as shown in (56) (Chomsky 1993 [1981a], p. 66):

(56) John tried [*PRO* to win]

According to Chomsky, *PRO* had to be regarded as a *pronominal anaphor*, as it shares characteristics with both pronominals and anaphors: on the one hand, «*PRO* is like overt pronouns in that it never has an antecedent within its clause or NP»; on the other, it «also resembles anaphors in that it has no intrinsic referential content [...], lacking specific reference» (Chomsky 1993 [1981a], p. 191). Given the fact that such an empty category would have to abide by both Principle A and B of Binding Theory, which in turn represents an contradictory conclusion, Chomsky's way to solve this problem was to assume that

«PRO has no governing category and is therefore ungoverned», thus satisfying Principle A and B vacuously (Chomsky 1993 [1981a], p. 191):

It is reasonable [...] to regard PRO as a pronominal anaphor. If so, it is subject to both the binding conditions (A) and (B). Then PRO is bound and free in its governing category, a contradiction if PRO has a governing category. Therefore PRO has no governing category and is therefore ungoverned. (*ibidem*)

Because of its property of being ungoverned, PRO «will be excluded from the complement positions governed by the head of some construction and from the position of subject of a tensed clause», the latter of which is governed by tensed INFL (Chomsky 1993 [1981a], p. 64). Moreover, from this property it also follows that, differently from pronouns, PRO «does not satisfy the Case Filter», as Case is assigned to an NP by means of government (Chomsky 1993 [1981a], p. 61).

Summing up the previous arguments, the obligatory occurrence of PRO in the non-case-marked subject position of infinitival clauses follows directly from its property of being ungoverned. However, this conclusion seems to be invalidated by the grammaticality of sentences such as (57), which shows that «phonetically-realized NP may appear as subject if the infinitive is in the context V-» (Chomsky 1993 [1981a], p. 66):

(57) I expect [him to leave]; I believe [him to be incompetent]

The fact that in (57) the lexical NP subject of the embedded infinitival does not trigger a violation of the Case Filter can be neatly explained by assuming that in English, as stated by Chomsky, in this specific context «Case can be assigned by the governing verb» as a marked option (*ibidem*). Given the exceptional nature of this phenomenon, which allows a verbal head to govern and assign accusative Case to the NP that follows it without being subcategorized for it, this property is still nowadays referred to as “Exceptional Case Marking” (ECM). Since «clausal complements are of the category S’, which we have assumed to be to be an absolute barrier to government», this intuition can be straightforwardly implemented by assuming that the infinitival clause in (57) is of the category S instead of S’ (*ibidem*). However, if not accompanied by further stipulations, this solution alone would in turn constitute a violation of the Projection Principle, an overarching constraint on the mapping between the different levels of representations:



(58) Projection Principle:

Representations at each syntactic level (i.e., LF, and D- and S-structure) are projected from the lexicon, in that they observe the subcategorization properties of lexical items. (Chomsky 1993 [1981a], p. 29)

This violation would follow from the fact that «by the projection principle, verbs with infinitival complements appear with clausal complements, as indicated by their lexical features» (Chomsky 1993 [1981a], p. 66). Accordingly, since the lexical feature of a clausal complement is S', any head selecting a clause – either tensed or untensed – as its complement subcategorizes for the type S', not for S; thus the former and not the latter should be selected as a complement at all levels of representation. At this point, Chomsky's solution to this problem was to assume that, in the so-called ECM constructions, the verbal head does actually subcategorize for an element S' but the S'-boundary is eliminated by a rule replacing the S'-boundary with an S-boundary, so that the embedded subject position is no more impervious to government/case marking:

A reasonable assumption [...] is that English has a marked rule of S'-deletion for complements of verbs of the *believe*-category, permitting the verb to govern the subject of the embedded complement, thus excluding PRO and permitting phonetically-realized NP in [57]. (*ibidem*)

In addition to differentiating unmarked cases like (56) from marked ones like (57), the existence of a marked, lexically-dependent rule deleting S' in infinitivals also accounted for the complementary distribution of a trace and a lexical subject in sentences like the following (Chomsky 1993 [1981a], p. 67):

(59) [<sub>NP</sub> e] was believed [COMP John INFL [<sub>VP</sub> be incompetent]]

(60) It was believed [<sub>S'</sub> that John is incompetent]

(61) John was believed [<sub>S</sub> t to be incompetent]

(59) represents the common D-structure to (60) and (61), with the only difference that INFL is tensed in the former but not in the latter. According to Chomsky's proposal, in the derivation of (60) *John* cannot be moved to the matrix subject position because, being S'-deletion allowed only for infinitival complements, it would not be able to bind its trace and there would therefore be a violation of Principle A of Binding Theory; for this reason, «*it*-insertion yielding [60] is obligatory» (*ibidem*). On the other hand, in the derivation of

(61) the application of Move  $\alpha$  to the embedded subject (which is crucially required because *believed*, being passive, is not a Case assigner and there would therefore be a violation of the Case Filter) is not blocked thanks to S'-deletion. Moreover, replacing S' with S also allows the subject trace to be governed despite INFL is untensed, thus satisfying the ECP (cf. Chomsky 1993 [1981a], p. 67).

Although «the option of deleting S' is in part a lexical idiosyncrasy», in Chomsky (1981a) the mechanism underlying ECM is described as an instance of parametric variation (Chomsky 1993 [1981a], p. 68):

These properties of PRO are not mere idiosyncrasies. We would like to reduce them, as far as possible, to general principles, even if these prove ultimately to be subject to some parametric variation, as in Exceptional Case-marking. (Chomsky 1993 [1981a], p. 71)

In conclusion, Chomsky's formulation of the lexical parameter of S'-deletion not only does represent an attempt to maintain a degree of descriptive adequacy in spite of the apparent violation of an universal principle of UG, as similarly done in Taraldsen's and Rizzi's pre-parametric inquiries, but it also fits well into within the explanatory model adopted by the P&P framework.

### 2.3 – Some theoretical remarks on the onset of the parametric enterprise

Despite the strong programmatic assumption put forth in Chomsky (1981a) in order to reconcile the idea of a principled UG with the existence of a system of parametric choices subsuming all possible non-peripheral cross-linguistic variation, the initial conception of the P&P model was neither particularly concerned with the locus nor the format of parametric variation. According to Rizzi, at this stage it was in fact «implicitly admitted that virtually any aspect of UG could be parametrized» (Rizzi 2013, p. 313). This holds particularly true for the very first formulation of the P&P in Chomsky (1981a), in which a certain degree of parametrization in each of the following subsystems of UG was hypothesised:

(62) The subsystems of principles of the theory of UG:

- i. bounding theory
- ii. government theory
- iii.  $\Theta$ -theory
- iv. binding theory
- v. Case theory
- vi. control theory

(Chomsky 1993 [1981a], p. 135)

In line with Rizzi's view, the approach followed in Chomsky (1981a) did not explicitly state which principles permitted a certain range of variation: it was simply suggested that «each of the systems outlined admits of some possibilities of parametric variation while certain principles are fixed» (Chomsky 1993 [1981a], p. 137). In all probability, the determining factor of the idea that (certain) principles could be parameterized was precisely Rizzi's *Violations of the Wh island constraint in Italian and the subjacency condition*, which was the first pre-parametric study built on the working hypothesis that parameters could be expressed on a universal constraint. The conclusion suggested by all these clues, therefore, would be that the original conception of *core grammar* regarded principles not as truly independent of parameters, but that the latter were assumed to correspond to structural options specified on the former (cf. Rizzi 2013, p. 313). This interdependence between principles and parameters was precisely what was regarded by generative linguists as the most attractive perspective when studying different languages. In fact, from that moment on, the general assumption in Generative Grammar has been that instances of linguistic variation would be no more regarded as an undesirable contingency hindering the study of the innate and universal properties of language faculty, but rather as an important means by which shedding light on those principles they are bound to:

[...] Study of closely related languages that differ in some clustering of properties is particularly valuable for the opportunities it affords to identify and clarify parameters of UG that permit a range of variation in the proposed principles. (Chomsky 1993 [1981a], p. 6)

Although it was not until the mid-eighties that the concept of parameter itself was refined in such a way that its possible locus and format was precisely defined, the further study of the very first parameters induced a small but important shift in the way parameters were related to principles. More precisely, an alternative hypothesis for the structure of

core grammar was that parameters are not directly expressed on universal principles, as previously assumed to hold without exception, but rather interact with their syntactic effects at a more superficial level. Reflecting this idea is the argument developed in Chomsky's *Principles and parameters in syntactic theory* (1981b), a short paper written in August 1979 which subsumes the conclusions drawn during the "Pisa lectures" of April 1979 and reinterprets them in the light of the most recent developments of parametric theory, building on comparative approaches to language variation. Although on a smaller scale, this work ideally starts from the point Chomsky's inquiry had been left after the recently concluded lectures, that is, from the confirmation that a truly comprehensive theory of UG needs to be flexible enough to describe all possible instances of cross-linguistic variation but, at the same time, has to be extremely rigid with respect to the innate principles constituting language faculty:

There is a tension between the tasks of description and explanation: to accommodate a wide range of phenomena, one is led to elaborate the descriptive devices, but to explain such phenomena, UG must be sharply restricted to only a few parameters in a highly constrained system. (Chomsky 1981b, p. 43)

At the time when this paper was published, the parameters that had been proposed in Generative Grammar were the parametrization of bounding nodes, the Null Subject Parameter and the lexical rule of S'-deletion. While these parameters did not receive any further substantial consideration with respect to their original proposals, they were briefly reviewed with special attention to the universal principles they were assumed to interact with. Regarding the parametrization of bounding nodes, Chomsky repropounded the classical comparison between the application of Move  $\alpha$  across a so-called wh-island in English and Italian, whose outcome is in turn represented by examples (63) and (64) respectively (Chomsky 1981b, p. 49):

- (63) [<sub>S'</sub> \**[To whom]*<sub>1</sub> [<sub>S</sub> did John wonder [<sub>S'</sub> *[which book]*<sub>2</sub> [<sub>S</sub> Mary had given *t<sub>2</sub> t<sub>1</sub>* ]]]]?
- (64) [<sub>S'</sub> *[A chi]*<sub>1</sub> [<sub>S</sub> John si domandò [<sub>S'</sub> *[che libro]*<sub>2</sub> [<sub>S</sub> Mary aveva dato *t<sub>2</sub> t<sub>1</sub>* ]]]]?

Assuming that in English S is a bounding node, in (63) the movement of *to whom* across two S nodes causes a violation of Subjacency, a universal principle which is

defined as «the property that an application of the rule “Move  $\alpha$ ” cannot cross more than a single bounding category» (Chomsky 1981b, p. 49). On the contrary, (64) shows that «in some languages, *eg* Italian, the analogues to such sentences as [63] are, however, grammatical» (Chomsky 1981b, p. 52). As pointed out by Chomsky, the impression that in Italian this principle can apparently be violated with impunity depends on the fact that in this language the bounding node which is relevant for Subjacency is not S, but S'. In these terms, while Subjacency is inviolable, on the other hand the notion of bounding category «is not determined by UG, but is rather left as an open parameter», and it is therefore the effect of this latter property of language which in turn yields all the possible different ways languages abide by this principle (Chomsky 1981b, p. 53).

In his comments about the Null Subject Parameter, Chomsky explicitly refers to the solution that would be shortly proposed by Rizzi regarding the possible account of the apparently inexplicable difference between pro-drop languages and non-pro-drop ones with respect to sentences like the following (Chomsky 1981b, p. 56):

(65) L'uomo che mi domando chi ha visto...

(66) The man that I wonder who saw...

As noted by Chomsky, the Italian sentence (65) «is actually ambiguous, with either the interpretation indicated in [67] or the interpretation with the indices on the traces interchanged» corresponding to [68] (*ibidem*):

(67) L'uomo [<sub>S'</sub> wh<sub>-1</sub> che mi domando [<sub>S'</sub> chi<sub>2</sub> t<sub>2</sub> ha visto t<sub>1</sub>]]...

“The man such that I wonder who saw him”...

(68) L'uomo [<sub>S'</sub> wh<sub>-1</sub> che mi domando [<sub>S'</sub> chi<sub>2</sub> t<sub>1</sub> ha visto t<sub>2</sub>]]...

“The man such that I wonder who he saw”...

In contrast, while for the English sentence (66) the former interpretation is possible, as shown in (69), on the other hand the latter interpretation, which is shown below as (70), is not (*ibidem*):

(69) The man [<sub>S'</sub> wh<sub>-1</sub> that I wonder [<sub>S'</sub> who<sub>2</sub> t<sub>2</sub> saw t<sub>1</sub>]]...

“The man such that I wonder who saw him”...

- (70) \*The man [<sub>S'</sub> wh-<sub>1</sub> that I wonder [<sub>S'</sub> who<sub>2</sub> t<sub>1</sub> saw t<sub>2</sub>]]...  
 "The man such that I wonder who saw him"...

As suggested by Chomsky, the factor which is responsible for the ungrammaticality of (70) is the ECP, since the empty category  $t_i$  in subject position cannot be properly governed by INFL, which in English cannot have pronominal status. At this point, however, the acceptability of the interpretation (68) in Italian seems to negate the fact that the same principle ruling out (70) cannot be violated. Chomsky's argument here is an anticipation of the problem pointed out and accounted for by Rizzi (1982) with respect to the apparent violation of the COMP-t effect in Italian wh-questions like *Chi credi che verrà* (see examples (53) and (54) above). And what interests Chomsky is precisely the fact that the possibility of free subject inversion in null subject languages creates a configuration which in turn allows the embedded subject to move to a higher position without violating the ECP:

There is [...] independent evidence that I will not review here that indicates that Italian does indeed observe this principle, to which we turn directly. Rizzi has proposed a possible solution to this dilemma. He presents evidence that the source of the competing interpretation to [67] is not the structure corresponding to [70] but rather one that involves prior inversion of the subject and verb, a process that is rather free in Italian. If this theory is correct, then the ambiguity of the sentence in question does not provide evidence against the inviolability of the principle governing [70] [...], contrary to appearances. (Chomsky 1981b, pp. 56-57)

As far as S'-deletion is concerned, the principle which is apparently at stake is the Case Filter. In this regard, the empirical data brought forth by Chomsky is based on the two-way distinction concerning verbs with infinitive complements, whose opposite poles are in turn represented by the verbs *try* (unmarked) and *believe* (marked/S'-deleting) (Chomsky 1981b, p. 61):

- (71) John tried [<sub>S'</sub>[<sub>S</sub> PRO to find Bill]]  
       \*John tried [<sub>S'</sub>[<sub>S</sub> Tom to find Bill]]  
 (72) \*John believed [<sub>S'</sub>[<sub>S</sub> PRO to be intelligent]]  
       John believed [<sub>S'</sub>[<sub>S</sub> Bill to be intelligent]]

The approach carried out here to the properties of S'-deletion is exactly the one which was proposed in Chomsky (1981a) in order to deal with the so-called ECM phenomenon. While the original argument will not be repeated here, the most important aspect pointed out in Chomsky's treatise is that, analogously to the relation between free subject inversion and the ECP in pro-drop languages, «while the Case Filter is inviolable, there is a parameter associated with it» which allows an apparent violation of this principle in conjunction with a specific parameter setting (which in this case is represented by the possible occurrence of verbs deleting S' and allowing ECM) (Chomsky 1981b, p. 61).

Considering the picture emerging from Chomsky's analysis of the parameters governing bounding categories selection, null subject licensing and S'-deletion, it can be argued that even if none of them is explicitly described as *directly specified* on the respective universal principle with which it interacts, nonetheless parametric variation is still assumed to interact with any UG principle dealing with the same grammatical aspect in defining what the possible outcome of syntactic derivation is. In these terms, the syntactic effects of Subjacency, ECP and the Case Filter on language, although unavoidable, are not the same cross-linguistically as they are conditioned by a finite set of optional grammatical properties varying from language to language:

To summarize so far, we have discussed three principles: subjacency, ECP, and the Case Filter. There are parameters associated with each of these principles: S may or may not be a bounding category for subjacency; a language may or may not have the inversion option of Italian, permitting apparent violations of ECP if it does; Verbs may or may not delete S', permitting apparent exceptions to the Case Filter if they do. (*ibidem*)

In conclusion, although only the Subjacency parameter had been explicitly formulated as a set of optional values specified on an invariant principle of UG, at the beginning of the P&P model also the Null Subject parameter and the parameter of S'-deletion were conceived as parameterized principles. In these terms, it can be argued that all the instances of parametric variation discovered so are regarded, at this theoretical stage, as the result of the interaction between universal principles and all possible (parametric) choices admitted by UG (cf. Chomsky 1981b, p. 71).

## 2.4 – Some notable systematic differences in the treatment of cross-linguistic variation

Within the early phase of GB-Theory, the core syntactic properties of each particular language were assumed to be determined by the different ways the values of the parameters associated with the universal principles of UG are fixed by the ideal speaker-hearer according to the set of well-formed sentences he or she is exposed to. In line with this assumption, which for a substantial part of the P&P framework was regarded as the new guiding principle in the research paradigm initiated by Chomsky nearly fifteen years earlier in the *Aspects* (1965), it comes as no surprise that most comparative studies carried out in Generative Grammar throughout the Eighties (including the ones presented in the previous paragraph) explicitly refer to the cases of cross-linguistic variation they aim at accounting for as syntactic parameters. However, even among the major works that contributed to the development of Chomskyan linguistics from a from a “principles-and-parameters” perspective, some of these studies do not expressly resort to the notion of parameter. From now on, special attention will therefore be paid not only to the specific contribution each of the following authors made to the P&P model, but also to whether this or that contribution had been originally formulated as either a syntactic parameter or a non-parametric systematic difference.

### *2.4.1 – Cross-linguistic variation at the syntax-semantics interface: a difference in the locus of wh-movement*

Regarding those instances of cross-linguistic differences which were not explicitly formulated as syntactic parameters, one of the most important ones in this regard is the variation concerning the locus of wh-movement which was proposed in Cheng-Teh James Huang’s PhD dissertation *Logical Relations in Chinese and the Theory of Grammar* (1982). This paper, which was one of the first works of the GB phase to focus on the syntax-semantics interface from a generative perspective, presents not only facts about the differences in the derivation of interrogative sentences between English and Chinese, but also some very influential insights into such different domains as phrase structure, scope relations between quantifiers and quantified expressions, configurationality, anaphora, movement constraints and the nature of Logical Form.



As far as the issue of wh-movement is concerned, Huang's analysis starts from the observation that, unlike in English, «the formation of wh questions, A-not-A questions» (more commonly labelled as *yes-no questions*) «and cleft sentences in Chinese does not involve the overt dislocation of any constituent» (Huang 1982, p. 19). Although this generalization does not specifically apply to interrogative sentences but can also describe the general behaviour of quantificational sentences, it is nonetheless true that the movement operation involved is not the same for both of these constructions. In fact, while on the one hand quantificational sentences involve covert movement in LF via QR to the S node, on the other hand interrogative sentences rely on wh-movement to the head COMP:

While all these constructions involve instances of Move  $\alpha$  in LF, we assume that quantificational sentences undergo May's QR, which performs adjunction, whereas wh questions, A-not-A questions, and clefts undergo abstract wh movement (to COMP), on a par with the overt wh movement process. (Huang 1982, p. 185).

In spite of such a difference in the syntactic movement distinguishing interrogative sentences from quantificational ones, the first point made by Huang is that wh-questions are actually quantificational in nature. This general idea, which had also been entertained by Chomsky (1975b), is then exemplified by means of the following sentences (Huang 1982, p. 251):

- (73) Everybody arrived.
- (74) [<sub>S</sub> [All x *such as* x a person] [<sub>S</sub> x arrived]]
- (75) Somebody arrived.
- (76) [<sub>S</sub> [Some x *such as* x a person] [<sub>S</sub> x arrived]]

The fact that the two LF representations (74) and (76) respectively represent the truth conditions of quantificational sentences (73) and (75) means that such sentences are nothing but «a shorthand [...] for all the sentences or propositions that satisfy the truth conditions that are related to each other by some logical connective, such as the conjunction or the disjunction» (*ibidem*). This is due to the fact that, as noted by Huang, «the structure [74] is a shorthand for the conjunction of all sentences which result from substituting a value for  $\underline{x}$  in the open clause [x arrived], i.e., conjunction of [Bill arrived], [John arrived], [Mary arrived], etc.», while a LF like (76) «is a disjunction of the instances of [x arrived], where x is a person», as either one value or another of multiple available

values can truthfully be substituted for x (Huang 1982, p. 251). Given such a view of the semantics of sentences (73) and (75), a wh-question such as (77) (Huang 1982, p. 252):

(77) Who arrived?

(78) [<sub>S</sub> [Which x *such as* x a person] [<sub>S</sub> x arrived]]

may be considered as a quantificational structure in which the wh-word is a logical operator binding a phonetically empty variable (as shown in (78)) but that, at the same time, presupposes the existence of a value, in turn represented by an existing specific individual, which can be truthfully substituted for x. In other words, question (77) «may be decomposed into the presupposition “Somebody arrived”», which corresponds to the disjunctive proposition asserted in (75), «and the request “Give me the identity of somebody”» (*ibidem*):

A natural way to look at a wh question [...] is to consider that it consists of a presupposition having the form of a quantificational sentence and a focus indicating the speaker’s request for a specification on the value of the quantificational element in the presupposition. (*ibidem*)

Accordingly, a very important consequence of the quantificational nature of wh-questions is that «wh-words in Chinese (as in English) are treated as scope bearing elements, on a par with ordinary Q-NPs» (Huang 1982, p. 254):

The scope bearing property of a wh word is a direct consequence of the fact that it corresponds to an indefinite quantificational expression (e.g., somebody) in the presupposition of a wh question. (Huang 1982, p. 255)

However, given the fact that in Chinese wh-movement is covert and thus not visible at S-structure, Huang’s analysis predicts that, in this language, wh-words may have a different scope interpretation from the one they apparently hold according to their overt position. This is particularly evident in the following examples (Huang 1982, p. 254):

(79) [Zhangsan wen wo [shei mai-le shu]]

ask I who buy-ASP book

“Zhangsan asked who bought books.”

- (80) [Zhangsan xiangxin [shei mai-le shu]]?  
           believe who buy-ASP book  
       “Who does Zhangsan believe bought books?”
- (81) [Zhangsan zhidao [shei mai-le shu]] (?)  
           know who buy-ASP book  
       a. “Who does Zhangsan know bought books?”  
       b. “Zhangsan knows who bought books.”

While the S-structures of sentences (79-81) only differ in their matrix verb, it is precisely the semantics of the verb itself which determines the scope of the quantifier expressed by the wh-word. In (79), the verb *wen* “ask” can only select a wh-question as its complement and therefore the wh-phrase takes scope over the embedded clause, yielding an indirect question. In (80), the verb *xiangxin* “believe” does not permit a wh-complement; consequently, the wh-word has wide scope and yields a direct question. Finally, in (81) the wh-word may take scope over either the embedded clause or the matrix one, as the verb *zhidao* “know” may optionally take an interrogative complement (cf. Huang 1982, pp. 254-255). According to Huang, the scope facts just pointed out for these three examples can be accounted for by postulating the existence, in Chinese grammar, of «an abstract wh movement rule» taking place at LF and creating the following LF representations (Huang 1982, p. 257):

- (82) [Zhangsan wen wo [[shei<sub>i</sub>] [ t<sub>i</sub> mai-le shu]]]  
           ask I who buy-ASP book
- (83) [[shei<sub>i</sub>] [Zhangsan xiangxin [ t<sub>i</sub> mai-le shu]]]  
           who believe buy-ASP book
- (84) a. [[shei<sub>i</sub>] [Zhangsan zhidao [ t<sub>i</sub> mai-le shu]]]  
           who know buy-ASP book  
       b. [Zhangsan zhidao [[shei<sub>i</sub>] [ t<sub>i</sub> mai-le shu]]]  
           know who buy-ASP book

Contrary to the scope interpretation of single wh-questions in a language like English, which is «readily given in the surface or S-structure representations of such questions», in a *wh-in-situ* language like Chinese the propositional function of the

interrogative phrase is not directly derivable from its surface position as the wh-quantifier covertly moves at LF (Huang 1982, pp. 256-257).

In order to strengthen the approach adopted so far, in his analysis Huang points out not only that wh-words, being equivalent to existential quantifiers, may undergo covert movement also in a language like English but also that, accordingly, the same scope ambiguities characterizing wh-questions in wh-in-situ languages may also emerge in wh-moving languages, although in contexts different from single wh-questions. A prime example of the possibility for a wh-word to undergo movement at LF in English is represented by the following multiple wh-question (Huang 1982, p. 261):

(85) Who bought what?

The situation associated with (85) is one in which the speaker already knows that someone bought something. Supposing, following Huang, that «someone  $x$  has extension ranging over the set of three individuals {John, Bill, Mary}, and that something  $y$  has extension ranging over the set of three things {the book, the pen, the pencil}», then this multiple wh-question expresses the presupposition that more than one person bought more than one thing (*ibidem*). However, differently from single wh-questions such as (78), what makes (85) relevant to Huang's argument is that the speaker «does not know the exact pairing between the instances of  $x$  and the instances of  $y$  and, by uttering [85], he requests information on the exact pairing», thus requesting an answer like (86) (*ibidem*):

(86) John bought the book, Bill bought the pencil, and Mary bought the pen.

According to Huang, the paired reading of (85) can be accounted for by assuming that, in multiple wh-question, the lower wh-word remains in situ at S-structure but, at the same time, it undergoes abstract wh-movement into COMP position. As a result of this process, «the wh word what is treated as a quantifier on a par with who, though only the latter occurs in quantifier position in SS» and, ultimately, both wh-words are treated as a single operator which binds two variables at the same time, as shown in the LF representation (87) (Huang 1982, pp. 261-262):

(87) [S' [COMP What<sub>i</sub> who<sub>i</sub> ] [S t<sub>i</sub> bought t<sub>j</sub> ] ] ?

For which pairing <x,y>, x a person and y a thing, x bought y.

As far as scope ambiguities showed by wh-words are concerned, one context allowing such phenomena in English is represented by the following case of multiple question interpretation, which had been previously investigated by Baker (1970) (Huang 1982, p. 262):

(88) [S' Who remembers [S' where we bought what ] ]?

The ambiguity of wh-interrogatives like this consists in the fact that (88) can be either answered as in (89a) or as in (89b). Considering (89a), (88) corresponds to «an ordinary direct inquiry on the identity of the matrix subject who», thus behaving like a single direct wh-question (*ibidem*). According to (89b), on the other hand, (88) is interpreted as «a direct inquiry on the pairing between who and the embedded unmoved what» and, therefore, has to be regarded as a multiple direct wh-question (Huang 1982, pp. 262-263):

(89) a. John does. John remembers where we bought what.

b. John remembers where we bought the book, Bill remembers where we bought the pencil, etc.

As anticipated above, the ambiguity between (89a) and (89b) can be explained by assuming that, «just as the wh words in Chinese, syntactically unmoved wh words in English also show scope ambiguities (though those already syntactically moved do not)» (Huang 1982, p. 262). More precisely, the account referred to by Huang consists in attributing this difference in interpretation to the two possible scopes of the wh-in-situ element *what*. In the interpretation of (88) requiring (89a) as an appropriate answer, the object *what* is covertly wh-moved into the embedded COMP position, thus pairing with the unmoved *where*. According to this reading, which in turn corresponds to the LF representation (90), the subject *who* is the only wh-element occurring in matrix COMP position and, therefore, the only one being assigned wide scope interpretation (Huang 1982, p. 263):

(90) [<sub>S'</sub> Who<sub>i</sub> [<sub>S</sub> t<sub>i</sub> remembers [<sub>S'</sub> [ what<sub>j</sub> where<sub>k</sub>] [<sub>S</sub> we bought t<sub>j</sub> t<sub>k</sub>]]]]?

In the other case mentioned, which corresponds to (91), the embedded object *what* undergoes abstract wh-movement into the matrix COMP position and is therefore assigned matrix scope together with the subject *who*. As a result, *what* and *who* are paired in a similar way as in (87) (Huang 1982, p. 263):

(91) [<sub>S'</sub> [What<sub>j</sub> who<sub>i</sub>] [<sub>S</sub> t<sub>i</sub> remembers [<sub>S'</sub> where<sub>k</sub> [<sub>S</sub> we bought t<sub>j</sub> t<sub>k</sub>]]]]?

Returning now to the preliminary generalization distinguishing quantificational sentences from wh questions, A-not-A questions and cleft sentences, the next step of Huang's argumentation is to account for the behaviour displayed by A-not-A questions with regards to covert wh-movement. An A-not-A question is a disjunctive, "yes/no" question whose appropriate answer has to identify either one of the two disjuncts introduced in the question itself, one being the exact opposite of the other. In Chinese, A-to-A questions can be formed by placing the negative morpheme *bu* to the left of a VP predicate, reduplicating the said VP (or a leftmost portion of it) and then placing the copy to the left of the negative (cf. Huang 1982, p. 277). Starting from a sentence like (92), the corresponding A-not-A question may take any of the forms in (93) depending on the portion of VP which is reduplicated (*ibidem*):

(92) ta xihuan ni.

"He likes you."

(93) a. [<sub>S</sub> ta [<sub>VP</sub> [<sub>VP</sub> xihuan ni] [<sub>VP</sub> bu xihuan ni]]]?

"Does he like you or not like you?"

b. [<sub>S</sub> ta [<sub>VP</sub> [<sub>V</sub> [<sub>V</sub> xihuan] [<sub>V</sub> bu xihuan]] ni]]?

"Does he like [you] or does [he] not like you?"

c. [<sub>S</sub> ta [<sub>VP</sub> [<sub>V</sub> [<sub>?</sub> [<sub>?</sub> xi-] [ bu xi]] -huan] ni]]?

"Does [he] or doesn't he like you?"

According to Huang, in Chinese «the [A-not-A] form is the result of some phonological rule of reduplication applying on the basis of some appropriate feature of modality [...] indicating the affirmative/negative modality», which in turn is referred to as the constituent AFF (Huang 1982, p. 282). If AFF is specified with the feature

[+affirmative], as in affirmative sentences, then it is phonologically null. On the other hand, in case it is [-affirmative], then it is overtly spelled out as the morpheme *bu*. However, what is postulated here is that the morpheme *bu* occurring in ordinary negative sentences is not the same as the one occurring in A-to-A questions. More precisely, in those contexts where it is not specified with either [+affirmative] or [-affirmative], AFF is [+Q], that is, a «quantificational ranging over [+affirmative] and [-affirmative]» (Huang 1982, p. 282). Differently from [-affirmative], this feature [+Q] does not imply the occurrence of the overt morpheme *bu* by itself, but rather by triggering the following rule (from this point on, Huang uses the feature [+A-not-A] instead of [+Q] for mnemonic purposes):

- (94) A-not-A Reduplication:  

$$[+A\text{-not-A}] [{}_{VP} X Y] \rightarrow [{}_{VP} [[X] [bu X]] Y]$$
not  
*(ibidem)*

Assuming with Huang that this rule takes place in PF, the S-structure of sentences (93a-c) before the application of A-not-A Reduplication is therefore represented by (95) (Huang 1982, p. 282):

- (95)  $[{}_S ta [+A\text{-not-A}] [{}_{VP} xihuan ni]]$   
he like you

Following the approach outlined for (75-78), as a consequence of its disjunctive nature «the [+A-not-A] constituent may then be interpreted as a quantifier ranging over the two members [A] and [Not A], i.e., [+affirmative] and [-affirmative]» (Huang 1982, p. 283). Because of its scope properties, the A-not-A constituent eventually undergoes movement at LF, thus yielding (88) (*ibidem*):

- (96)  $[[+A\text{-not-A}]_i [ta t_i xihuan ni]]$   
he like you

If used in A-not-A questions, according to Huang «an [+A-not-A] constituent may be [+wh]» (*ibidem*). The immediate consequence of such featural specification is that the covert movement undergone by the [A-not-A] constituent in the derivation of A-not-A questions in Chinese is not QR but wh-movement. The LF representation (96) can thus be

converted into the semantic interpretation (97), which contains a *wh*-operator binding a trace (Huang 1982, p. 283):

- (97) [[For which x such as x ∈ {[+A], [-A]}] [s ta x xihuan ni]]  
he like you

The examples used by Huang in order to support the account just outlined show that, in Chinese, A-not-A questions behave in an analogous manner to wh-questions with regards to both their quantificational nature and covert movement properties. First, exactly on a par with the examples of wh-questions showed in (79-81), A-not-A questions may have different scope interpretation from the one suggested by the position of the [+A-not-A] constituent at S-structure. In the following examples (98), (99) and (100), [lai-bu-lai] has embedded scope, matrix scope and ambiguous scope respectively (Huang 1982, p. 284):

- (98) Wo xiang-zhidao [Lisi lai-bu-lai]  
I wonder come-not-come  
“I wonder whether Lisi will come or not.”
- (99) Ta shuo [Lisi lai-bu-lai] ?  
He say come-not-come  
“Did he say that Lisi will come, or did he say that Lisi won’t come?”
- (100) Ta zhidao [Lisi lai-bu-lai] (?)  
He know come-not-come  
a. “He knows whether Lisi will come or not.”  
b. “Does he know that Lisi will come, or does he know that Lisi won’t come?”

Another example proposed by Huang concerns the kind of movement undergone by the [+A-not-A] constituent in sentences other than wh-questions. If the [+A-not-A] constituent occurs in the scope position of a quantification scope marker at S-structure (that is, before the application of A-not-A Reduplication), then it would be interpreted as a quantifier rather than a wh-element, thus undergoing QR instead of wh-movement. This scenario is shown by example (101), as the [A-not-A] constituent occurs in an embedded subject sentence which, in turn, is within the scope of the universal quantification marker *dou* “all” (Huang 1982, p. 283):



- (101) [s [s ta [+A-not-A] xihuan ni] dou meiyou guanxi]  
 like you all no matter  
 “Whether or not he likes you, it doesn’t matter.”

As shown by (102), the LF representation of (101) reflects the quantifier interpretation of the [+A-not-A] constituent, which in this specific context behaves like a universal quantifier in both being adjoined to the matrix S via QR (contrary to being moved into matrix COMP) and its semantic interpretation (Huang 1982, p. 283):

- (102) a. [s [+A-not-A]<sub>i</sub>] [s [s ta t<sub>i</sub> xihuan ni] dou meiyou guanxi]]  
he like you all no matter  
b. [[For both x such as x ∈ {[+A], [-A]}] [[ta x xihuan ni] dou meiyou guanxi]]  
he like you all no matter

After accounting for wh-questions and A-not-A questions, the final step of Huang's argumentation is to account for the behaviour displayed by cleft sentences. In Chinese, cleft sentences take the following form (Huang 1982, p. 289):

- (103) Zhangsan shi mingtian dao Niuyue qu.  
           be tomorrow to N.Y. go  
       ‘It is tomorrow that Zhangsan will go to New York.’

As showed by (103), in Chinese clefts the copula *shi* is immediately followed by the focus of the sentence, which in this specific case is the phrase [tomorrow], while the rest of the sentence is taken as the presupposition (cf. *ibidem*). As noted by Huang, in the English translation of (103) there is overt movement of the focus phrase from its base position within the presupposed clause as a result of its relativization, as shown by (104) (*ibidem*):

- (104) It is tomorrow<sub>i</sub> [that Zhangsan will go to New York t<sub>i</sub>]

However, Huang points out that in Chinese «a cleft sentence differs from a non-cleft only in the presence vs. absence of the focus indicator, the copula shi» (Huang 1982, p. 290). In fact, according to Huang, the process of cleft sentence formation can simply be

regarded as the insertion of the element *shi* immediately in front of the constituent in focus of a non-cleft sentence as the following (Huang 1982, p. 290):

- (105) Zhangsan mingtian dao Niuyue qu.  
tomorrow to N.Y. go  
“Zhangsan will go to New York tomorrow.”

To summarize so far, while in English cleft sentences involve overt movement of the focus in syntax, in Chinese such constructions «involve no overt dislocation of the focus in Syntax and consequently, unlike their English counterparts, do not exhibit a quantifier-variable configuration at SS» (*ibidem*).

At this point, one important aspect which is pointed out in Huang’s treatment is that, despite being as well focalizing constructions, pseudo-cleft sentences cannot be accounted for by means of the approach adopted for cleft sentences analysis (Huang 1982, p. 291):

- (106) [wo mingtian yao mai de] shi [neiben shu]  
I tomorrow want buy DE be that book  
“What I want to buy tomorrow is that book.”

According to Huang, in constructions like (106) «what appears after the copula shi is the pseudo-clefted focus, and what appears before it is a free relative whose empty head is coreferential with the focus» (Huang 1982, p. 292). This amounts to saying that, unlike what happens in the case of cleft sentences, the derivation of pseudo-clefts involves overt movement of the focus from its base position in the relative, as shown by the S-structure representation of (106) shown in (107) (*ibidem*):

- (107) [<sub>NP</sub> [<sub>S</sub> wo mingtian yao mai t<sub>i</sub>] de [e]<sub>i</sub>] shi [neiben shu]<sub>i</sub>  
I tomorrow want buy DE be that book

Coming back to cleft sentences in Chinese, Huang proposes that the correct way to account for their properties is to assume that *shi* is a focus operator dominated by the element EMP indicating the emphatic modality and, therefore, to consider it «as an adverb, on a par with modals, negation, etc.» (Huang 1982, pp. 292-293). In these terms, what

distinguishes a cleft sentence from a simple sentence in Chinese is simply the fact that the former, but not the latter, features the EMP element. Therefore, contrary to the derivation of pseudo-clefts in the same language and clefts in English, for Chinese clefts there is no need for postulating neither overt movement nor any antecedent-gap or quantifier-variable relation involved at S-structure (cf. Huang 1982, p. 293). Regarding their LF representation, Huang's proposal is that «in LF, a maximal phrase immediately following the focus operator is incorporated into EMP, and is treated as the focused constituent» (*ibidem*). This is exemplified by the following cleft sentence, which has the S-structure (108) and the "first stage" LF representation (109) (*ibidem*):

- (108) wo shi mingtian yao mai neiben shu  
 I FO tomorrow want buy that book  
 "It is tomorrow that I want to buy that book."  
 (109) [<sub>S</sub> wo [<sub>EMP</sub> shi mingtian] [<sub>VP</sub> yao mai neiben shu]]  
 I FO tomorrow want buy that book

After the incorporation stage shown in (109), the phrase dominated by EMP moves to COMP via wh-movement and gives, although at LF rather than SS, a quantifier-variable relation analogous to that occurring in English cleft sentences after the relativization of the sentence's focus. As a result, «like the English counterpart and like the pseudo-clefts in both languages, the cleft sentence in Chinese is also structurally dichotomized into the two portions focus and presupposition» (Huang 1982, pp. 293-294):

- (110) [<sub>S</sub> [<sub>EMP</sub> shi mingtian]<sub>i</sub> [<sub>S</sub> wo t<sub>i</sub> yao mai neiben shu]]  
 FO tomorrow I want buy that book

Similarly to what happens with quantified expressions and wh-phrases, the latter of which include the [+A-not-A] constituent found in A-not-A questions, what makes it possible for the focused constituent in a cleft sentence in Chinese to undergo movement at LF is, according to Huang, its quantificational nature. This intuition is in turn based on the proposal, first formulated in Chomsky (1976), that the rule of semantic interpretation determining the focus of a sentence attributes to the focused element the status of a bound variable (cf. Chomsky 1977b [1976], p. 203). Thus, given sentence (111), the rule FOCUS applied to the NP *John* gives the LF representation (112) (Huang 1982, p. 294):

(111) His mother loves JOHN.

(112) [[For x=John] [his mother loves x]]

As previously hinted with regards to cleft sentences, Huang's intuition is that a sentence containing a focused element corresponds to a presupposition which, in turn, has the form of a quantificational sentence. In the case at hand, (111) can be decomposed into the presupposition (113), in which the focally stressed constituent *John* corresponds to the quantificational expression *someone*:

(113) His mother loves someone.

[[For some x such that x a person] [his mother loves x]]

As far as cleft sentences are concerned, a crucial consequence of the quantificational nature of the focused element is that, on a par with quantified expressions, wh-phases and A-not-A phrases, the phrase immediately following *shi* may exhibit scope ambiguities. This fact is particularly evident when comparing the behaviour of clefted constituents with that of proper names (Huang 1982, p. 299):

(114) Zhangsan xiangxin [Lisi mingian lai]

believe tomorrow come

“Zhangsan believes that Lisi will come tomorrow.”

One of the properties of definite NPs like *Zhangsan* and *Lisi* is that, as proper names, they «usually take the widest possible scope in an utterance» (Huang 1982, p. 297). Accordingly, in a sentence like (114), *Lisi* can only take wide scope – that is, only over the matrix clause, thus yielding only one interpretation (for the sake of clarity, its interpretation is equivalent to (116a) below). By contrast, in a sentence like (115), a focused proper name may have either narrow or wide scope (Huang 1982, p. 298):

(115) [Zhangsan xiangxin [shi Lisi mingian lai]]

believe FO tomorrow come

As observed by Huang, «the clefted constituents following shi do not always take wide scope even though they are clefted definite NPs [...] exactly as wh phrases, A-not-A

phrases, and Q-NPs» (Huang 1982, p. 298). For example, in (115) *shi Lisi* does behave like a quantificational element rather than a proper name in being allowed to take scope over either the entire sentence or the embedded clause. This causes the ambiguity of (115) between (116a) and (116b) (*ibidem*):

- (116) a. It is  $Lisi_i$  that Zhangsan believes  $t_i$  will come tomorrow.  
 b. Zhangsan believes that it is  $Lisi_i$  that  $t_i$  will come tomorrow.

The two interpretations of (115) differ according to whether the belief that there is someone who will come tomorrow is held by the speaker of the entire sentence or by the matrix subject. According to (116a), it is the speaker who presupposes that someone will arrive tomorrow, asserting that such person is *Lisi*. According to (116a), on the other hand, it is *Zhangsan* that holds this belief. The possibility of two distinct interpretations for (115) can be therefore regarded as a direct consequence of the fact that the focused element may undergo wh-movement at LF to either the matrix or the embedded COMP position, «exactly as in the case of wh words or A-not-A constituents embedded under an optional interrogative matrix verb like “know” or “remember”» (*ibidem*).

Although throughout Huang (1982) the term *parameter* is occasionally resorted to in order to account for some specific typological differences like, for example, that intercurring between configurational languages and non-configurational/scrambling ones, this notion is not even mentioned with regards to the overt vs covert wh-movement opposition outlined above. Nevertheless, the fact that modern generative literature refers to the latter distinction as a parameter (see for example Rizzi 2014, p. 18) suggests that Huang’s analysis of wh-movement in Chinese has all the requirements – although only *a posteriori* – to be treated as such. This conclusion can be ideally elicited by means of at least two points. First, similarly to the other parameters reviewed so far in this chapter, according to Huang’s analysis the identical behaviour of wh-questions, A-not-A questions and cleft sentences in Chinese with respect to both surface order and semantic interpretation can be accounted for by a single grammatical property located at a deeper level, that is, the quantificational nature of wh-phrases and focused constituents. Second, in line with the assumption that parameters correspond to structural options specified on universal principles, Huang states that although on the one hand wh-movement is a universal linguistic property in its own right, on the other hand individual grammars may differ as to which level of representation this syntactic operation applies. In fact, while in

English *wh*-movement is placed between D-structure and S-structure, in Chinese it is placed between S-structure and LF – that is, at the same locus as QR in both languages (cf. Huang 1982, p. 186):

A typological view inherent in this way of looking at questions in languages like Chinese is that language families do not differ with respect to whether they have a *wh* movement rule or not; rather, all languages are assumed to incorporate such a rule as a substantive universal, but may differ in where they use the rule, in Syntax or in LF. (Huang 1982, p. 254).

#### 2.4.2 – Preposition stranding and V-P reanalysis

Among the cross-linguistic differences which were noted and analyzed in Generative Grammar during the first part of the Eighties, some notable examples include the ones which are discussed in the fifth chapter of Richard Kayne's work *Connectedness and Binary Branching* (1983). In this chapter, Kayne begins his treatment by outlining two systematic differences in the syntactic properties of English and French which he believes to be directly related to each other. First, while English allows preposition stranding (P-stranding) in *wh*-constructions and in passives, French does not (Kayne 1983, p. 103):

- (117) a. Which candidate have you voted for?  
b. \*Quel candidat as-tu voté pour?
- (118) a. John was voted against by almost everybody.  
b. \*Jean a été voté contre par presque tous.

Second, while English allows epistemic verbs of the *believe*-class to be merged with infinitive complements featuring a lexical subject, French does not (*ibidem*):

- (119) John believes Bill to have lied.
- (120) \*Jean croit Bill avoir menti.

According to Kayne, the nature of the relation between P-stranding and ECM constructions can effectively be clarified by looking at the differences intercurring between the complementizer systems of English and French. Starting from the assumption that «English has the complementizers *that* and *for*», with French *que* being the direct

counterpart of English *that*, the first step in this direction is to verify whether *de* in French effectively corresponds to English *for* (Kayne 1983, pp. 103-104). As noted by Kayne, such an assumption is entirely reasonable if considering the following examples. First, similarly to *que* in (121), also *de* in (122) seems to be analyzable as a complementizer at the beginning of the sentential complement introduced by the matrix verb (Kayne 1983, p. 104):

- (121) Je lui ai dit qu'il parte.  
"I told him that he (should) leave."  
(122) Je lui ai dit de partir.

Another piece of evidence presented by Kayne concerns the fact that, being part of a sentential complement but not of a nominal complement, *de* cannot occur in (132) (*ibidem*):

- (123) a. Je lui ai dit quelque chose.  
"I told him something."  
b. \*Je lui ai dit de quelque chose.

At this point, one possible objection would be that, in the cases mentioned so far, *de* is not a complementizer but rather corresponds to the English infinitival *to*, which in turn introduces an infinitive complement S (IP). In this regard, however, the fact that in the translation of (124) *to* is part of the indirect wh-question while, as shown in (125), *de* cannot occur in the same position contributes to dispel this doubt (*ibidem*):

- (124) Je lui ai dit où aller.  
"I told him where (to) go."  
(125) \*Je lui ai dit où d'aller.

Furthermore, the complementizer nature of *de* directly accounts for the ungrammaticality of (125), which on these terms can be assumed to be a consequence of the violation of the doubly-filled COMP filter. This general restriction excludes the co-occurrence of a wh-phrase and a complementizer in a COMP-position, thus ruling out (125) as well as English sentences like (126) (Kayne 1983, p. 105):

(126) \*I told him where for her to go.

Assuming that French *de* is in fact the counterpart of English *for*, Kayne argues for a parallelism between French *de* and Italian *di* by means of extending to the latter the data considered so far for the former. Like French *de*, Italian *di* can be analyzed as part of a sentential complement (as in (127)) but not as part of a non-sentential complement (as in (128)), and it cannot co-occur in COMP-position with a *wh*-phrase without violating the doubly-filled COMP filter (as shown in (123-124)) (Kayne 1983, p. 105):

- (127) Gli ho detto di partire. (= [121])  
him (I) told leave  
“I told him to leave.”
- (128) a. Gli ho detto qualcosa. (= [123a])  
b. \*Gli ho detto di qualcosa. (= [123b])
- (129) Gli ho detto dove andare. (= [124])
- (130) \*Gli ho detto dove di andare. (= [125])

Another interesting aspect noted by Kayne is that French *de* and Italian *di* behave in a similar way also regarding control. In fact, both *de* and *di* occur in a variety of control contexts, involving many cases of object and subject control (*ibidem*):

- (131) Je lui ai interdit/suggéré/demandé de partir.  
 "I prohibited/suggested (to)/asked him to leave."  
 (132) Gli ho proibito/suggestito/chiesto di partire.

While being useful for further setting apart *de* and *di* from *for*, the data just outlined above concerning control structures in French and Italian brings to the table some aspects which will eventually allow Kayne to propose his account of the differences in P-stranding and ECM constructions between French and English. In these terms, while «*de* and *for* now have in common their status as complementizers, in particular as infinitival complementizers», there is an important difference between them. On the one hand, although *de* cannot introduce an infinitival clause featuring a lexical subject, *for* can (Kayne 1983, p. 109):



(133) \*Ce serait dommage de quelque chose lui arriver.

(134) It would be a pity for something to happen to him.

On the other hand, *de* is compatible with control verbs selecting infinitival complements with a null/PRO subject, while *for* is not (Kayne 1983, p. 109):

(135) Ce serait dommage de partir maintenant.

(136) \*It would be a pity for to leave now.

The reason behind the ungrammaticality of (133) with respect to (134) has to be attributed, as noted by Kayne, to a violation of the Case Filter. In fact, it is the lexical subject *quelque chose* that triggers such a violation due to its impossibility to receive Case from either the embedded infinitival INFL node or the complementizer *de*. Following the same reasoning, the grammaticality of (134) derives from the fact that, contrary from its French counterpart, *for* can assign Case to the embedded lexical subject (cf. Kayne 1983, pp. 109-110).

Turning now to the ungrammatical status of (136), Kayne interprets this piece of data in the light of the general principles of Case Assignment, as first formulated in Chomsky (1980):

(137) General principles of Case Assignment:

A. NP is oblique when governed by P and certain marked verbs;

B. NP is objective when governed by V;

C. NP is nominative when governed by Tense;

(Chomsky 1980, p. 25)

Starting from the assumption that Case is assigned under government, according to Kayne «it follows also that *for* must govern the adjacent subject position» in both (132) (134) (Kayne 1983, p. 110). This idea is then interpreted in the light the definition of PRO as a pronominal anaphor as proposed in Chomsky (1981a), according to which PRO can only occur in an ungoverned syntactic position. This reasoning, together with the assumption that *for* is a Case assigner, leads Kayne to the conclusion that the subject position immediately following *for* is governed. Hence, it is government of the adjacent subject position by *for* which, being incompatible with control, rules out the PRO in (122) (cf. Kayne 1983, p. 110).

As far as the grammaticality of (135) is concerned, the approach adopted by Kayne directly follows the account proposed for (136). Since *de* cannot assign Case, then it is reasonable to assume that «*de* in COMP does not govern the adjacent subject position» (Kayne 1983, p. 110). It then follows that in such context control is possible, and this explains the well-formedness of (135).

From what has been argued so far regarding the distinct behaviour of *for* and *de* with respect to government, Kayne proceeds to generalize the behaviour of these two prepositional complementizers to all the other elements of the same kind in English and French respectively:

The promised intermediate generalization is a slight one: English prepositional complementizers govern the adjacent infinitival subject position, but French prepositional complementizers do not. (*ibidem*)

This point is very important, as it is precisely by means of this generalization that Kayne aims at accounting for the difference between *croire* in French and *believe* in English. In fact, Kayne's idea is that the effects of this generalization can crucially be seen on both prepositional complementizers and epistemic verbs of the *believe*-class. First, Kayne observes that the preposition *à*, which in French can also function as a complementizer, does indeed confirm the above generalization as «it is compatible with control and incompatible with a following lexical subject» (*ibidem*):

- (138) \*Je cherche quelqu'un à Jean photographe.  
I look for someone for/to John (to) photograph.  
(139) \*Marie est facile à Jean contenter.  
Mary is easy for/to John (to) please.

Second, while French verbs of the *believe*-class do not allow an embedded NP subject in postverbal position, wh-extraction applied to the embedded subject in the very same construction yields a well-formed interrogative sentence, as shown by the contrast between (140) and (141) (Kayne 1983, p. 111):

- (140) \*Je crois/reconnais/constate Jean être le plus intelligent de tous.  
"I believe/acknowledge/have determined John to be the most intelligent of all."

(141) Quel garçon crois/reconnais/constates-tu être le plus intelligent de tous?

“Which boy do you believe/acknowledge/have determined (to) be the most intelligent of all?”

This particular property sharply distinguishes French verbs like *croire* from all those verbs which, rather than being subcategorized for an element S', directly select a NP complement, as «in V NP PP or V NP NP or V NP AP or simply V NP, it is never the case, as far as we know, that the postverbal NP must be moved by *Wh* Movement to be licit» (Kayne 1983, p. 111). The different grammaticality status between (140) and (141) can then be accounted for by assuming, in the spirit of Chomsky (1981a), that the syntactic structure of both these sentences corresponds to V [<sub>S</sub> NP VP]. In these terms, (140) is ungrammatical because *Jean*, being the lexical subject of an infinitive, cannot receive Case from within the infinitival S itself: this amounts to saying that (140) is ruled out by the Case Filter. On the other hand, (141) is grammatical because the *wh*-movement of *quel garçon* to the matrix subject position allows the former to receive Case from the matrix verb, thus avoiding violation of the Case Filter (cf. Kayne 1983, pp. 111-112).

Although the analysis outlined above for *croire* is not substantially different from the one suggested in Chomsky (1981a) for unmarked, non-ECM verbs exemplified by *try*, here is where the similarity between the two models ends. As noted by Kayne, the fact that the English word-for-word counterpart of (140) is grammatical without the need for the postverbal NP to move suggests that «our firm conclusion that the French equivalents of *believe*, etc., take an infinitival S' complement (whose lexical subject cannot remain in place because of Case considerations) does not automatically transpose to English» (Kayne 1983, p. 112). On the one hand, Chomsky (1981a) proposed that, differently from unmarked verbs, the matrix verb *believe* can govern the subject position of the infinitive by means of a lexically-dependent rule allowing ECM. However, on the other hand, Kayne does not regard the resort to ECM as a satisfactory account for the difference between English and French with regards to epistemic verbs of the *believe*-category; rather, what he considers to be crucially relevant in the framework of Chomsky's *Lectures* in this respect is the mutually exclusive relation between government and control. Such negative relation, which was shown to differentiate prepositional complementizers *for* and *de* in (133-136), holds true also of the class of verbs at issue since control is excluded in English but not in French, as shown in the following examples (*ibidem*):

(142) \*I believe/acknowledge/affirm to have made a mistake.

(143) Je crois/reconnais/affirme avoir fait une erreur.

The reason why Kayne does not accept Chomsky's idea of an irreducible, lexical-based difference between the two languages is that, although valid from a descriptive point of view, an account solely based on the lexically-dependent rule of S'-deletion does not aim to adequately explain anything about the relation between the *believe/croire* opposition and the different properties of prepositional complementizers in English and French:

Let us now ask the following question: why do French and English differ here as they do? And let us take the position that even Chomsky's answer is not satisfactory. That is, let us ask why *believe* should allow government of the embedded subject position, but not *croire*. Why is it not the other way around? Our answer will exploit the similarity, within the Case/government framework, between the *believe/croire* difference and the *for/de* difference [...]. (Kayne 1983, p. 113)

Starting from the assumption that there is no S'-deletion rule allowing *believe*-type verbs to govern/assign Case to the embedded subject position via ECM, the fact that «although V can govern an NP in COMP, i.e. across a single S-type (S,S') boundary, it cannot govern across two S-type boundaries» implies that, according to Kayne, in English sentences like (119) the matrix verb does not actually govern the subject of the infinitive (Kayne 1983, p. 112). Instead, Kayne's proposal is that the latter is governed and assigned Case by a null prepositional complementizer which is in turn selected by the former:

In particular, let us assume [...] that *believe*-type verbs take a  $\Phi$  complementizer. We assume further that  $\Phi$  is another prepositional complementizer, which differs from *for* and *de* in having no phonetic realization. (Kayne 1983, p. 113)

The null hypothesis implied by Kayne is that verbs of the *believe*-category take a  $\Phi$  complementizer in both English and French, as «there is no reason why French should not have a prepositional  $\Phi$  complementizer with the same class of verbs» (*ibidem*). Since in these terms the subject of the infinitive receives its Case from  $\Phi$  rather than the matrix verb, the apparent difference between *believe* and *croire* actually stands in the behaviour of the prepositional complementizers by which they are respectively followed, and which in

turn conforms to the generalization distinguishing English prepositional complementizers from French ones with regards to government:

From this point of view, there is no essential difference between *believe* and *croire*. The apparent differences between them are rather a function of the way in which English and French treat prepositional complementizers. (Kayne 1983, p. 113)

The contrasts between (119) and (120) and between (142) and (143) (here repeated as (144-145) and (146-147)) can thus be accounted for without postulating either an essential difference between the verbs *believe* and *croire* or the existence of a lexical ECM rule deleting the upper clausal boundary S' from V [<sub>S'</sub> NP VP]. According to this approach, the difference in the grammaticality status between (144) and (145) follows from the fact that  $\Phi$  governs *Bill* in (144) but not in (145), with the consequence that only in the former case the embedded subject can be assigned Case and, therefore, prevent a violation of the Case Filter:

(144) John believes Bill to have lied.

(145) \*Jean croit Bill avoir menti.

Regarding the contrast between (146) and (147), in (147) subject control is allowed because  $\Phi$ , though present, does not govern the embedded subject position. Conversely, given the fact that in English prepositional complementizers govern the embedded subject position,  $\Phi$  blocks control in (146), hence its ungrammaticality:

(146) \*I believe/acknowledge/affirm to have made a mistake.

(147) Je crois/reconnais/affirme avoir fait une erreur.

The final steps of Kayne's analysis bring out the relation between the *believe/croire* difference – which, as seen above, is argued to derive in turn from the difference between English and French prepositional complementizers – and the difference in P-stranding between English and French. By assuming «the existence, in English only, of a Reanalysis rule that amalgamates V and P into one constituent», Kayne proposes not only that the possibility of stranding a preposition depends on the availability of V-P reanalysis, but also that such specific rule is in turn directly conditioned by the governing properties of prepositions (Kayne 1983, p. 114). It is now worth mentioning that the effect of reanalysis

on preposition stranding had been first put forth in Hornstein & Weinberg (1981), which also proposed the original formulation of this syntactic rule:

There is a general syntactic rule of Reanalysis which says that in the domain of VP, a V and any set of contiguous elements to its right can form a complex V. The rule would be roughly of the form [...]  $V \rightarrow V^*$  (where V c-commands all elements in  $V^*$ ). (Hornstein & Weinberg 1981, p. 60)

The idea of a specific link between V-P reanalysis and P-stranding constructions seems to be confirmed by the fact that a non-preposition stranding language like French does indeed lack reanalysis between preposition and verb. Moreover, it is important to observe that the unavailability of P-stranding in French does not follow from an absolute absence of reanalysis constructions, but from the specific impossibility of applying such rule to a verb and a preposition. In fact, as Kayne notes, French has reanalysis of V-V, as in causatives, «but apparently no reanalysis of V (X) P, contrary to English» (Kayne 1983, p. 115):

Consequently, rather than interpreting the lack of preposition stranding in French as resulting from the absence of a reanalysis rule, let us state more precisely that it results from the absence of a reanalysis rule involving prepositions. (*ibidem*)

The fact that, in French, reanalysis of V-V is allowed while V-P reanalysis is not suggests that, according to Kayne, «there must be some important difference between V and P at issue» (*ibidem*). In these terms, Kayne's proposal is that «reanalysis between two lexical categories is possible only if the two govern in the same way» and that, in the matter in question, the difference in P-stranding between English and French depends on the fact that, while in English both verbs and prepositions govern structurally the element they subcategorize for, in French only verbs do (Kayne 1983, p. 116):

In French, P and V do not govern in the same way; but in English they do. (That is, in English, P can govern structurally, as well). (*ibidem*)

This basic difference between verbs and prepositions conforms to the assumption that, in the sense of Chomsky (1981a), verbs generally assign *structural* objective *Case* by means of the structural relation of government, while prepositions in French (and other languages not allowing P-stranding) assign *inherent* oblique *Case* by means of their

inherent properties of subcategorization/argument selection (cf. Chomsky 1993 [1981a], pp. 170-171).

In these terms, the opposition between English and French with regards to P-stranding is a result of the interaction of V-P reanalysis and the ECP. By assuming, along the lines of Hornstein & Weinberg (1981), that the category P is never a proper governor, in a language like English V-P reanalysis generates as its output a complex verb which acts as a proper governor for the object of P. Accordingly, the trace left by wh-movement in (117a) (here repeated as (148) before and after reanalysis) is properly governed and thus conforms to the ECP:

- (148) a. You have [<sub>V</sub> voted] [<sub>PP</sub> for which candidate]  
b. You have [<sub>VP</sub> [<sub>v</sub> [<sub>V</sub> voted] [<sub>P</sub> for]] which candidate]  
c. [Which candidate]<sub>i</sub> have you [<sub>VP</sub> [<sub>v</sub> [<sub>V</sub> voted] [<sub>P</sub> for]] <sub>t<sub>i</sub></sub>?

In languages where V-P reanalysis is not allowed, on the other hand, the complement of P is not properly governed and therefore the ECP is violated. This is exactly what happens in the French example (117b) (here repeated as (149)):

- (149) \*[Quel candidat]<sub>i</sub> as-tu voté [<sub>PP</sub> pour <sub>t<sub>i</sub></sub>]?

Given this difference between prepositions and verbs in French, in the configuration [{V/P} [<sub>S</sub> NP X]] the embedded NP subject can only be assigned Case by V, but not by P. In fact, while the domain of head-government by the verb – that is, the domain of structural Case assignment – extends to the nearest S-boundary, as in (141), the government domain of the prepositional complementizer is restricted to its sister node due to subcategorization, with the result that «there are no such instances of cross-S' Case Assignment from P» (Kayne 1983, p. 115). Another example put forth by Kayne concerns the French complementizer *de*. When occurring in the configuration P [<sub>S</sub> NP X], as in (133), *de* cannot assign Case to the NP since it is not subcategorized for the subject of S. On the other hand, its prepositional status – and the consequent impossibility to govern the embedded subject – ensures its compatibility with control, as shown by the well-formedness of (135) (cf. Kayne 1983, pp. 115-116).

In conclusion, what emerges from Kayne's account is that both the absence of two seemingly independent properties like P-stranding and ECM constructions in French and

their co-occurrence in English derive from an individual difference in the governing properties of prepositions. This abstract difference between French and English not only appears to attain adequacy at both descriptive and explanatory level, but it is also compatible with, although without explicitly referring to, the P&P model of cross-linguistic variation outlined in Chomsky (1981a).

#### 2.4.3 – *The complementary distribution of finite verb and lexical complementizer in verb-second languages*

Another seminal comparative work not relying on the notion of parameter is *On the Interaction of Root Transformations and Lexical Deletive Rules* (1983) by Hans Den Besten. In this paper, whose first version already circulated as a manuscript in 1977 and was published four years later, Den Besten aimed at overcoming the typology of transformations which had been set forth in Emonds (1976) and, more precisely, the sharp distinction between *root transformations* and *structure-preserving transformations* – namely, those which were assumed to apply exclusively on root sentences and those which were applicable also on non-root sentences respectively – as only the latter type had the additional requirement of moving constituents only to positions which had already been base-generated by phrase structure rules. As Den Besten (1981) deals with a set of theoretical issues which are not directly relevant to the aim of the present work, attention is therefore focused on the 1983 Appendix which Den Besten later added to his original paper. This section starts from the assumption that, as observed throughout the entire paper, in languages such as Dutch and German lexical complementizers and verb movement to complementizer position are in complementary distribution:

First, note that the Verb Preposing rules I discuss in this paper without exception induce an obligatory rule of Complementizer Deletion. This complementary distribution of preposed finite verbs and lexical complementizers gives one the impression that Verb Preposing (SAI [Subject-Auxiliary Inversion], Subject-Clitic  $\bar{V}$  Inversion) substitutes the finite verb for COMP. (Den Besten 1989 [1983], p. 88)

This generalization is shown in the following German example. While on the one hand the lexical complementizer *daß* may co-occur with the finite verb *komme* when the



latter is in clause-final position, as in (150a), when the finite verb moves up to COMP there is no overt complementizer, as in (150b) (Den Besten 1989 [1983], p. 82):

- (150) a. Er sagte, daß er morgen komme. (Bach & Horn 1976)  
           He said, that he tomorrow comes (subjunctive)  
       b. Er sagte, er komme morgen.  
           He said, he comes (subj.) tomorrow

The same is true for Dutch. As shown in (151), the complementizer *of*, while overt when the finite verb is not moved from its base-generated, clause-final position, is not phonologically realized when the verb is fronted to COMP position (Den Besten 1989 [1983], pp. 23-24):

- (151) a. --, of je broer nog komt.  
           --, whether your brother yet comes  
       b. Komt je broer nog?  
           Comes your brother yet

According to Den Besten, the above data can be accounted for by adopting for V-to-C movement, a canonical root transformation, a similar analysis to the one that was assumed to hold for structure-preserving transformations such as *wh*-movement, the reason being that «root transformations share with the cyclic rule of *Wh*-Movement the property of being Complementizer Attraction Rules» – namely, the property of moving a syntactic element to a position immediately dominated by a matrix *S'* (Den Besten 1989 [1983], p. 89). In these terms, Den Besten's proposal is that all Complementizer Attraction Rules can be formalized as structure-preserving, regardless of whether they are root or non-root. Their common rule schema is the following (*ibidem*):

- (152) Complementizer Attraction

X	-	[+F <sub>i</sub> ]	-	Y	-	[c+F <sub>i</sub> ]	-	Z
1		2		3		4		5
1		4		3		e		5

Since in (152) «C is some constituent, and  $F_i$  is some morphosyntactic feature», the rule schema of any Complementizer Attraction Transformation is obtained by substituting  $[+F_i]$  with the morphosyntactic feature specified on both the moved constituent and its landing site (Den Besten 1989 [1983], p. 89). As far as wh-movement is concerned, the landing site  $[+WH]$  of the fronted constituent is base-generated outside COMP, which Den Besten indicates as  $[\pm T]$  in the following expansion rule (*ibidem*):

$$(153) S' \rightarrow [+WH] [\pm T] S$$

Together with the phrase structure rule in (153), the rule schema in (152) describes the fact that, in Dutch (but also German) wh-questions, the fronted wh-phrase is moved higher than COMP. Moreover, assuming that V-to-C movement occurs in wh-questions in both Dutch and German, it is clear that in Dutch example (154b) the  $[\pm T]$  position is filled with the finite verb, which moving from its base position in turn hinders the occurrence of the otherwise obligatory lexical complementizer, as in (154a) (Den Besten 1989 [1983], pp. 23-24):

- (154) a. --, welk boek (of) hij wil lezen.  
           --, which book (whether) he wants read  
       b. Welk boek wil hij lezen?  
           Which book wants he read

Returning to the generalization regarding the complementary distribution of lexical complementizers and Verb Preposing, Den Besten's proposal is that in Verb Preposing languages COMP is a  $[\pm Tense]$  position. As pointed out by the author, this analysis of C as an inflectional category is supported by the fact that specific lexemes belonging to this syntactic category are associated with specific classes of verb forms, although generally not with specific tenses (cf. Den Besten 1989 [1983], p. 90). This is in turn confirmed by the behaviour of Dutch lexical complementizers, as «*dat* "that" and *of* "whether, if" are  $[+T]$  complementizers and *om* "for" [...] requires a *te*-infinitive» (*ibidem*). By applying this intuition to the general rule schema in (152) and, therefore, assuming that the morphological feature  $[F_i]$  specified on the landing site of Complementizer Attraction Rules is also specified on the syntactic element the said rule applies, Den Besten formulates Verb Preposing as Move Tense (*ibidem*):

(155) Move Tense (Verb Preposing)

X	-	[+T]	-	Y	-	[v+T]	-	Z
1		2		3		4		5
1		4		3		e		5

As Den Besten previously hinted at, according to (155) Verb Preposing «substitutes the finite verb for COMP» (Den Besten 1989 [1983], p. 88). As a consequence, the mutually exclusive relationship between Verb Preposing and the presence of a lexical complementizer derives from the fact that the COMP node can only be filled by one element, either the overt complementizer or the finite verb. In the former case, as COMP is already filled, the finite verb has no position to move to and therefore stays in its base position. Conversely, in all the cases when COMP is not occupied by a complementizer, the inflectional nature of C triggers application of Verb Preposing, hence moving [<sub>v</sub>+T] to [+T] position:

This new formalization of the rule of Verb Preposing predicts that there will be Verb Preposing only if the corresponding lexical complementizer is absent - since the fronted finite verb occupies the complementizer position – and that there may be a lexical complementizer if the verb is not moved (modulo other rules such as *Wh*-Movement which may influence the presence of a complementizer). (Den Besten 1989 [1983], p. 90)

Although the comparative analysis carried out in Den Besten (1983) is focused exclusively on West Germanic languages, which are all verb-second languages with the only exception of English, and thus does not in principle take into account non-verb-fronting grammars, Den Besten's proposal neatly fits into the general P&P perspective. In fact, the idea that Verb Preposing directly depends first and foremost on the presence of a [+Tense] feature in COMP, and then on the absence of a lexical complementizer in the same position, suggests that the fundamental difference between languages such as Dutch and German on the one side and languages such as English on the other can be attributed to the fact that, while in the former there is a [+Tense] feature in COMP which must be lexically realized, in the latter C is not an inflectional category and thus there is no need for V-to-C movement.

## 2.5 – The main parameters of the Government-Binding Theory from the mid-Eighties onwards

### 2.5.1 – The parameterization of the Projection Principle

Returning to the main syntactic parameters explicitly formulated during the GB phase, another important example in this sense is the parameterization of the Projection Principle proposed in Hale (1983). In this paper, whose title is *Warlpiri and the grammar of non-configurational languages*, Hale's declared aim is to find a parameter which could be held responsible for some of the major differences between English and Warlpiri, an Aboriginal language of Central Australia. As noted by Hale, the importance of comparing English and Warlpiri lies in their typological distance, as these two languages ideally represent "canonical" exemplars of the *configurational* and *non-configurational* types (cf. Hale (1983), p. 5). According to the syntactic model put forth in Chomsky (1981a), configurational languages are those languages in which «grammatical functions (GFs) such as subject-of, object-of, and so on [...] are determined in terms of syntactic configurations» (Chomsky 1993 [1981a], p. 42). On the opposite pole, in non-configurational languages «GFs are not represented in D- and S-structures [...] in terms of the formal structures, but are assigned randomly to D-structures and [...] to S-structures» (Chomsky 1993 [1981a], p. 132).

In spite of the formal definitions given above, the thing in which Hale appears to be most interested in the paper at issue is the cluster of superficially visible properties which distinguish non-configurational languages as Warlpiri from configurational languages like English:

- (156) a. free word order
- b. syntactically discontinuous semantic expressions
- c. extensive use of null anaphora

The fact that in Hale (1979) the expressions «“free word order” or “scrambling” languages» referred to the linguistic type to which Warlpiri was held to belong hints at the fact that free word order is possibly the most striking feature shared by those languages to which the label “non-configurational” has been traditionally applied (Hale 1981 [1979], p. 1). Free word order in Warlpiri is exemplified by the fact that, with the only restriction that the auxiliary element (AUX) must be in second position, sentences containing the same

content words in permuted orders are perceived by Warlpiri native speakers as repetitions of one another, as shown by the following example (Hale 1983, p. 6):

- (157) a. Ngarrka-ngku ka wawirri panti- rni.  
Man      ERG AUX Kangaroo spear NONPAST  
“The man is spearing the kangaroo.”

Hence also:

- b. Wawirri ka panti-rni ngarrka-ngku.
- c. Panti-rni ka ngarrka-ngku wawirri.
- d. Ngarrka-ngku ka panti-rni wawirri.
- e. Panti-rni ka wawirri ngarrka-ngku.
- f. Wawirri ka ngarrka-ngku pantirni.

As far as property (156b) is concerned, if in configurational languages the constituents of a noun phrase must appear in a position linearly adjacent to one another, in Warlpiri the elements forming a single semantic expression may be discontinuous at surface level. This is shown in (158) (*ibidem*):

- (158) Wawirri kapi-rna panti- rni                   yalumpu.  
kangaroo AUX      spear NONPAST that  
“I will spear that kangaroo.”

As noted by Hale, the position of the AUX element in the example above indicates that *wawirri* has to be regarded as a single syntactic constituent. Nonetheless, the interpretation of (158) is exactly the same as (159), in which the elements *wawirri* and *yalumpu* form the single syntactic constituent corresponding to “that kangaroo” (*ibidem*):

- (159) Wawirri yalumpu kapi-rna panti- rni.  
kangaroo that      AUX      spear NONPAST

Last, the term “null anaphora” refers to the fact that a non-configurational language does frequently display cases in which arguments are not realized overtly. This property

may apply to the object, as in (160), to the subject, as in (161), and to both, as in (162) (Hale 1983, p. 7):

- (160) Ngarrka-ngku ka panti- rni.  
 man      ERG AUX spear NONPAST  
 "The man is spearing him/her/it."
- (161) Wawirri ka panti- rni.  
 kangaroo AUX spear NONPAST  
 "He/she is spearing the kangaroo."
- (162) Panti- rni              ka.  
 spear NONPAST AUX  
 "He/she is spearing him/her/it."

The next step undertaken by Hale for explaining the examples proposed so far is to define the essential features of Warlpiri's phrase structure by means of the following rewriting rules (*ibidem*):

- (163) a.  $X' \rightarrow X'^* X$   
 b.  $V' \rightarrow \text{AUX } X'^* V X'^*$

The first X-bar schemata refers to nominal expressions and infinitive clauses and specifies that the head of these constituents must occur in final position. The second rule defines the phrasal structure of finite clauses by expressing two requirements: first, their head V do not need to occur in final position; second, finite clauses must have an auxiliary (although it is generated in first position, this element can subsequently be inserted into second position depending on other factors of phonological nature. Moreover, the obligatory co-occurrence of the AUX element with the verbal head of a tensed clause derives from the fact that the former can actually be regarded as a part of the latter, hence the possibility for the verbal element to be overtly realized as a discontinuous syntactic entity (cf. Hale 1983, pp. 7-8).

Returning now to the superficial characteristics of non-configurational languages listed in (150), Hale describes them as direct consequences of the permissiveness which Warlpiri's phrasal structure is allowed by the X-bar schemata of (163). One of the crucial factors in determining this remarkable freedom with respect to languages belonging to the configurational type is that, in the two rewriting rules seen above, «the symbol X

designates a categorially vacuous node» which is not specified for any categorial feature and, therefore, does not impose any restriction to lexical insertion (Hale 1983, p. 7).

Under such assumptions, the possibility of free word order in Warlpiri derives from the fact that, given that every symbol  $X'$  in (163) can be substituted by any categorial feature, lexical insertion is free to insert any type of constituent in any linear order. Furthermore, the existence in Warlpiri of both discontinuous nominal expressions and extensive null anaphora can be seen, although only in part, as a consequence of the system outlined above (cf. Hale 1983, pp. 8-9). Starting from the assumption that lexical insertion is free for every unspecified  $X$  in the representation of  $V'$  in (163b), the occurrence of discontinuous nominal expressions can be assumed to be, although only in part, the result of the fact that «any nominal lexical item can be inserted at any  $X$ » (Hale 1983, p. 9). Similarly, the possibility of null anaphora for all arguments in Warlpiri can be partially accounted for by means of the format adopted in Hale's notation, as any number of  $X'$  constituents, including none, is allowed to appear in the set of strings generated by Warlpiri's phrasal structure rules in conjunction with the starred symbol  $X'^*$  (cf. *ibidem*).

Although the approach followed so far provides a straightforward description of Warlpiri's three most striking typological properties, an account solely based on phrase structure rules like (163) cannot, as pointed out by Hale himself, do the same with respect to explanatory adequacy. According to the original model of Generative Grammar, that is, the one laid out in Chomsky's *Aspects of the Theory of Syntax* (1965) and which had been referred to as *the standard theory*, the set of context-free rewrite rules forming the categorial component of the base was widely assumed to encode both hierarchical structure and linear order. Ever since the shift from standard theory to the EST phase in the early Seventies, however, «much recent work in theoretical linguistics [...] suggest[ed] very strongly that certain (perhaps most) aspects of phrase structures are derivative of independent grammatical processes and principles» (Hale 1983, p. 10). A most notable example of such works was Chomsky's seminal paper *Remarks on Nominalization* (1970). Not only did this paper introduce the X-bar schema of phrase structure, but also contained the proposal that lexical heads carry a piece of information known as *subcategorization frame* – namely, a set of categorial features which specify the number and type of the syntactic arguments with which the lexical item needs to co-occur (cf. Chomsky 1993 [1981a], pp. 35-36). It goes without saying that, in these terms, the existence of categorial rules turned out to be an unbearable redundancy within the overall structure of grammar,

as the same information these rules would specify about any phrase would also be encoded in lexical heads of the latter (cf. Stowell 1981, p. 71):

The X-bar theory of phrase structure, for example, itself incorporates as a fundamental principle the notion of that “phrasal types” (i.e., phrase internal “levels” of structure, symbolized by numbers of bars or primes) are projections of lexical categories (cf. Chomsky, 1970; Jackendoff, 1977; George, 1980) suggesting that the hierarchical dimension of phrase structure is basically a lexical matter, rather than a matter pertaining to an autonomous phrase structure component. Moreover, the constituency of phrases is, to a very large extent, predictable from the subcategorizational properties of their lexical heads (cf. Grimshaw, 1981; Bresnan, 1982), suggesting a further diminution in the role of phrase structure as an autonomous system. (Hale 1983, p. 10)

Assuming, therefore, that the information expressed by phrase structure rules is in fact redundant with respect to the information already encoded by strict subcategorization frames at lexical level, Hale did not choose either to base the non-configurational characteristics of Warlpiri on the properties of the categorial component or to take the latter out of the picture altogether. Instead, he approached the problem from another direction and postulated the existence of two parallel levels of representations both projected from the lexicon: *lexical structure* (LS) and *phrase structure* (PS):

In the light of these considerations, I would like now to look at the matter from another angle. Specifically, I would like to explore the possibility that the typological distinction at issue here finds its origins not in phrase structure itself but, rather, in the nature of the relationship between phrase structure (PS) and LEXICAL STRUCTURE (LS). (Hale 1983, p. 11)

According to this “parallel model” of syntactic representation, the LS consists in the argument structure of a predicate, while the PS corresponds to the syntactic representation. The LS is specified by each verbal lexical entry along with other information such as its categorial designation, its phonological form and its dictionary definition. For example, *panti-rni*, a verb of the agent-patient semantic class whose meaning is “to pierce, to poke, to jab, to spear” can be said to have the following LS, where *arg<sub>x</sub>* corresponds to the subject and *arg<sub>y</sub>* to the object respectively (Hale 1983, p. 12):

(164) [*arg<sub>x</sub>*, *arg<sub>y</sub>*, *panti-rni*]



For verbs belonging to the agent-patient semantic class, the subject/agent is assigned ergative Case and the object/patient is assigned absolutive Case, as shown in (157). Other verbal classes display other Case associations. For example, for monadic verbs like *wangka-rni* “to speak” the subject is associated with the absolutive Case (Hale 1983, p. 13):

- (165) Kurdu ka    wangka- rni.  
           child AUX speak    NONPAST  
           “The child is crying.”

According to Hale, the associations between thematic roles and their respective Case categories is specified on the predicate’s LS itself. Thus, the LS of a verb like *panti-rni* can be formalized as follows (Hale 1983, p. 14):

- (166) [*erg*<sub>x</sub>, *abs*<sub>y</sub>, panti-rni]

The relationship between the two concepts of LS and PS is introduced precisely with regards to these Case labels associated with the arguments in LS. In fact, in Hale (1983) the grammatical function of these labels is to establish, by means of the following rule, a link between the Case of each nominal constituent in PS and the Case of each argument position in LS:

- (167) Linking Rule:  
       Co-index N’ in PS with *arg* in LS, provided the case category of N’ is identical to that of *arg* (assigning a distinct index to each *arg* in LS). (*ibidem*)

Although the Linking Rule guarantees identity between nominal expressions in the syntactic representation and the predicate’s thematic roles, «there is nothing in the rule itself, or in the grammar of Warlpiri generally, which prevents the linking of more than one N’ in PS to a single argument in LS», as showed by all the instances of syntactically discontinuous expressions occurring in Warlpiri (Hale 1983, p. 15). Having said that, one condition this rule is subject to is that, according to Hale, an N’ linked by rule (167) to a argument in the LS of a verb must occur in a PS position which is sister to that verb. Considering the fact that the auxiliary can be regarded as a part of the verb, then the

Linking Rule and this latter condition jointly account for the grammaticality of Warlpiri sentences featuring discontinuous constituents, as shown in (159) (cf. Hale 1983, p. 8).

As far as the extensive use of null anaphora is concerned, the occurrence of this phenomenon can be accounted for by assuming that, in Warlpiri, while on the one hand nominal constituents in PS must be properly linked to arguments in LS, «the dependency is not reciprocal» (Hale 1983, p. 16). Hence, the possibility for argument positions encoded by the predicate to lack their syntactic counterpart(s) at PS, as shown by sentences (160-162).

After these premises, which aimed at briefly pointing out the mechanisms behind the mappings between phrasal constituents and lexical arguments in Warlpiri, the next step of Hale's argumentation was to provide evidence of the fact that the latter are structured in a hierarchical way at LS level. In these terms, Hale's proposal concerns the possibility of attributing «a configurational structure to LS, in place of the "flat linear" structure depicted heretofore (e.g., [166])» by resorting to the idea of a structural asymmetry between the argument corresponding to the subject and the other arguments of the verb, with particular reference with the direct object (Hale 1983, p. 22). The diagnostic chosen here in order to bring out the existence of this asymmetry in Warlpiri consists in testing the behaviour of some cases of anaphoric and pronominal binding with respect to principles A and B of Binding Theory. The approach adopted by Hale is straightforward but ingenious: given the fact that both anaphors and pronominals must entertain an asymmetrical c-command relation with their respective antecedents, by assuming that the LS of a verb corresponds to the governing category of the arguments of the said verb it can be thus verified whether this level of representation features an internal structure by looking at the behaviour of the binding relations between the subject and the object occurring in Warlpiri's tensed sentences:

Simplifying matters somewhat, suppose we say, further, that the LS of a tensed clause constitutes the governing category of each argument which it contains. This will force an anaphor to be bound within LS, and it will force a non-anaphor to be free therein. Now, assuming that an anaphor cannot c-command its antecedent (cf. Reinhart, 1976), we can account for the unidirectional character of the binding relation involved in the reflexive-reciprocal construction if we assume that LS, rather than having the flat structure depicted in [166] above, has an internal syntactic organization over which an asymmetrical c-command relation can be defined. (Hale 1983, p. 23)

As an illustration of the relation between Binding Theory and the nature of LS representation in Warlpiri, the relevant cases set forth in this sense consist in some instances of the reflexive-reciprocal construction, an example of which is provided below (Hale 1983, p. 21):

- (168) Kurdu- jarra- rlu ka- pala- **nyanu** paka- rni.  
 Child dual ERG PRES 33subj **refl** strike NONPAST  
 “The two children are striking themselves/each other.”

The most important element to look at in (168) is the auxiliary *ka-pala-nyanu*. In fact, if the verb is dyadic, the AUX element encodes all the information concerning the person and the number of both the subject and the object. These pieces of information, which are realized morphologically as overt person markers, appear in a precise order immediately after the auxiliary base: first, there is the person marker *subj* construed with the ergative argument; second, there is the person marker *obj* construed with the absolutive argument (cf. Hale 1983, p. 17). In case the sentence requires a reflexive-reciprocal interpretation, «the *obj* marker **-nyanu** (glossed *refl*) occurs in place of the ordinary *obj* person markers», thus indicating that the object argument in LS is anaphorically bound to the subject (Hale 1983, p. 21). Crucially, while on the one hand Hale’s data show that an object can be marked with *-nyanu*, on the other hand «there is no alternative morphology according to which the *subj* person marker, rather than *obj*, is replaced by a special form indicating that the subject is bound within its governing category (e.g., by the object)» (Hale 1983, p. 23). Consequently, Hale’s conclusion is that the subject argument in the LS of a tensed clause cannot be marked as anaphoric because of the existence, in the architecture of LS, of a structural difference between arguments according to which the subject asymmetrically c-commands the object, but not the reverse. This representation of Warlpiri’s argument structure at LS can therefore be seen to conform to argument selection in configurational languages, since the manner in which different thematic roles like agent and patient – and hence grammatical functions like subject and object – are assigned in languages of this latter type is defined in terms of positions in a configurational syntactic structure.

Assuming at this point that the configurational nature of the structure of LS in Warlpiri, a language of the non-configurational type, assuredly proves the null hypothesis that «all languages are configurational at that level of linguistic representation», Hale

proposes that the typological difference between configurational and non-configurational languages can be plausibly attributed to the distinct way the level of lexical representation can be mirrored by phrase structure (Hale 1983, p. 25):

Perhaps, then, the place to look for the fundamental difference between configurational and nonconfigurational languages, as these labels are commonly used, is in the relation between LS and PS – i.e., specifically, in the manner in which particular languages instantiate the principle according to which properties of lexical items are projected onto syntax. (*ibidem*)

Since according to the GB framework the mapping between each level of representation is fulfilled by the Projection Principle, in order to account for the fact that «some languages mirror LS configurationality in PS, while others do not», Hale proposes that this principle has to be parameterized (*ibidem*). Note that the Projection Principle assumed by Hale (1983) consists in a restricted version of the one originally formulated in Chomsky (1981a). The reason behind this reformulation is that Hale is specifically interested in the argument array that the verb selects at LS – that is, in more general terms, in its subcategorization properties:

(169) The Projection Principle (restricted):

If *verb* selects *arg* at  $L_i$ , then *verb* selects *arg* at  $L_j$  (where  $L_i, L_j$  range over the “levels” LF, D-structure, S-structure in the syntactic representations of clauses).

(*ibidem*)

Here there is the core of Hale’s typological proposal, which is stated in the form of his *Configurationality Parameter*:

(170) The Configurationality Parameter (CP):

a. In configurational languages, the projection principle holds of the pair (LS, PS).

b. In non-configurational languages, the projection principle holds of LS alone.

(Hale 1983, p. 26)

According to this parameter, in configurational languages (Hale’s revised version of) the Projection Principle holds of both LS and PS, so that there is a relation of identity

between the arguments of a verb in LS and the syntactic constituents in PS in terms of structural positions:

The projection principle, as it is enacted in configurational languages, determines a fixed and uniform relation between noun phrases occupying specific positions in PS and arguments occupying the corresponding positions in LS. (Hale 1983, pp. 29-30)

By contrast, in non-configurational languages «the CP does not determine any connection at all between LS and PS» and, as a result, the argument structure of a verb is not required to be mirrored at PS (Hale 1983, p. 27). This means not only that PS does not need to be hierarchically structured, but also that the distribution of the verb's arguments at this level of representation is entirely optional – with the only exception, in Warlpiri, of the conditions depending on the Linking Rule (167), which imposes each nominal expressions at PS to be associated with the matching argument at LS but not the other way around). Hence, *ceteris paribus*, the co-occurrence of such typological properties as free word order, the possibility of discontinuous expressions and the extensive use of null anaphora.

In addition to providing an account for the linguistic properties listed in (156), another advantage of the approach followed by Hale (1983) lies in the fact that the conception of configurationality which follows from the parameterization of the Projection Principle correctly predicts certain other characteristics distinguishing Warlpiri from configurational languages such as English. For example, while English has transformational rules of the type referred to as NP-movement, Warlpiri lacks them. Since NP-movement is to be seen as movement from a theta-marked argument-position to a non-theta-marked argument-position, the impossibility for non-configurational languages to have NP-movement rules follows from the fact that, in languages of this type, «there are, strictly speaking, no argument positions in PS, where the category NP (or N') is instantiated» (Hale 1983, p. 28):

In non-configurational languages, [...] the term "argument" is appropriate only in reference to the *arg* terms in LS, in the initial functional structures of clauses at least. (Hale 1983, p. 30)

Another example in this regard is the absence, in non-configurational languages, of certain categories of constituents like the base-generated empty elements [<sub>NP</sub> e] occupying the subject position of passives at D-structure level and the PRO occurring in control

clauses. Given the fact that the appearance of both [<sub>NP</sub> e] and PRO at PS/D-Structure level of configurational languages is forced by the Projection Principle, from this it follows that such categories will be lacking in non-configurational languages as this principle is assumed to not hold of PS in languages of the latter type (cf. Hale 1983, p. 29).

The non-configurational setting of the CP has also repercussions on the way Case marking functions in Warlpiri with respect to languages like English. Starting from the assumption that the Linking Rule implicitly requires that «the “abstract” case (i.e., that assigned in LS) must agree with the morphological case (i.e., that assigned to a nominal in the lexicon)», in a configurational language morphologically overt Case assignment will appear to mirror the Case associations already specified for the arguments in the verbal LS, the reason being that the corresponding setting of the CF determines a relation of identity between PS and LS (*ibidem*). The opposite is true, however, in a non-configurational language like Warlpiri. Therefore, no correlation between the structural position of a nominal in PS and its morphological Case is necessarily required in languages of the latter type.

In the concluding remarks of his paper, Hale briefly presents some important implications of the conception of configurationality emerging from the proposed relationship between LS, PS and the Projection Principle. More precisely, Hale’s aim here is to suggest the existence of some further parametric options which, he argues, would follow from the parameterization of the Projection Principle:

If the CP corresponds, in fact, to a true parameter of linguistic variation, then it generates certain corollary parameters of its own, particularly for languages of the non-configurational type. Since the CP itself does not determine any particular relation between LS and PS in non-configurational languages, there is a large potential for variation among languages in the manner in which these entities relate to general principles of grammar. (Hale 1983, p. 42)

Although the CP represents the core of Hale’s (1983) account of the fundamental difference between configurational and non-configurational languages, the ensuing argumentation really makes clear not only what the implications of assuming a structureless PS architecture for non-configurational languages are but also, accordingly, for which reason the presence of underlying hierarchical structure at D- and S-structure was assumed be subject to cross-linguistic variation at the time that Hale proposed the parameterization of the Projection Principle.

Returning to Hale's work, the hypothesis put forth here is that languages of the non-configurational type may show a certain degree of variation according to which structure is relevant to the application of principle C of Binding Theory. In these terms, according to Hale, «there could be three non-configurational subtypes», each one of which would be generated by a distinct parameterization of this principle (Hale 1983, p. 44):

- (171) a. Only PS is relevant:  
 Surface word order is free, *ceteris paribus*, except that Binding Condition (C) cannot be violated in PS.
- b. Only LS is relevant:  
 Word order is free, *ceteris paribus*, but Condition (C) cannot be violated in LS.
- c. Both LS and PS are relevant:  
 Word order is free, *ceteris paribus*, except that Condition (C) cannot be violated (in LS or PS).  
 (Hale 1983, pp. 44-45)

As stated above, this tripartition is only a proposal needing to be backed up by further empirical data (especially regarding languages of the types (171a) and (171b), which Hale set forth because of conceptual necessity but without actually providing any relevant linguistic example). Having said that, however, the type (171c) seems to perfectly apply to Japanese – a language that, belonging to the non-configurational type, is regarded by Hale as being «possessed of a flat PS structure» (Hale 1983, p. 44). The “flat” nature of Japanese's PS structure directly follows from the setting (b) of the CP (170) and, crucially, constitutes the very same factor determining the lack of NP-movement rules in Warlpiri. Since in a purely linear PS the direct arguments of a verb symmetrically c-command each other, it follows that «in a flat structure» like the one at issue «both [...] precedence and c-command [...] are relevant to Condition (C)» and, therefore, a sentence in which a subject R-expression is co-indexed with an anaphoric object will not be ruled out if these two constituents appear in this order (*ibidem*). This hypothesis seems to be confirmed by the following data from Japanese (*ibidem*):

- (172) John<sub>i</sub>- ga    zibun<sub>i</sub>- o    hihan    si- ta. (Whitman 1982)  
 NOM self    ACC criticize do PAST  
 “John criticized himself.”

Conversely, principle C will rule out the very same coreference chain if the anaphoric object occurs before the subject R-expression in the linear order (Hale 1983, p. 44):

- (173) \*Zibun<sub>i</sub>- o John<sub>i</sub>- ga hihan si- ta. (Whitman 1982)  
 self ACC NOM criticize do PAST  
 “Himself, John criticized.”

As further evidence of this dual binding requirement to R-expressions, Hale finally sets forth an example in which, although the anaphor precedes the R-expression binding it, the latter is not c-commanded by the former. In the particular case of (174), what prevents the anaphor *zibun* from c-commanding *John* although preceding it is the fact that, as noted by Hale, the former constituent is within the object NP, thus excluding mutual c-command between them in PS (*ibidem*):

- (174) Zibun<sub>i</sub>- no haha- o John<sub>i</sub>- wa aisite i- ru (Whitman 1982)  
 self GEN mother ACC THEME loving be NONPAST  
 “His own mother, John loves.”

If, on the one hand, (172-174) attest that PS is relevant to principle C in Japanese, on the other hand in these three examples the R-expression c-commands the anaphor in LSs, hence the impossibility of determining whether LS is also relevant. This latter question, however, is answered affirmatively by means of the following example, in which «*John* and *kare* “he” cannot be coreferential, even though coreference in PS would not violate Condition (C)» (Hale 1983, p. 45):

- (175) \*John<sub>i</sub>- no sensei- o kare<sub>i</sub>- ga syookai si- ta (Saito 1982)  
 GEN teacher ACC he NOM introduce do PAST  
 “John’s teacher, he introduced.”

Crucially, the reason of the ungrammaticality of (175) is not to be found in PS, but rather in LS. As noted by Hale, in fact, in LS the subject R-expression *John* is c-commanded by the anaphoric subject *kare*, hence the violation of principle C of Binding Theory (*ibidem*).



(176) [<sub>V</sub> kare-ga = *arg*<sub>x</sub> [<sub>V</sub> John-no sensei-o = *arg*<sub>y</sub>, syookai si-ta]]

After examining the symmetrical c-command relation between arguments in Japanese and some of its implications on the interpretation of sentences, it is interesting to see how in Hale (1983) the issue of formulating a syntactic parameter constraining the projection of underlying structures arises not only to account for Warlpiri's core non-configurational properties but also, more in general, in order to find a unique and discrete solution to their common source – namely, the uniform lack of syntactic structure attributed to languages of the non-configurational type. In fact, in this work the question of non-configurationality is approached as an actual problem in need of a verifiable solution, the reason being that the very existence of languages such as Warlpiri and Japanese would otherwise run counter to the idea of a non-parameterized Projection Principle, which in principle excludes the possibility of non-configurational languages.

In conclusion, the Configurationality Parameter as formulated in Hale (1983) offers a principled explanation for both the set of observed differences distinguishing Warlpiri from configurational languages and, on a deeper level, the possibility itself of having some exceptions with respect to the universal validity of the Projection Principle. In these terms, Hale's proposal does not differ from the ones made by other linguists who, like Taraldsen and Rizzi, aimed at preserving the descriptive adequacy of the generative model by parameterizing the effects of a universal principle of UG.

### 2.5.2 – The parameter of abstract incorporation

If on the one hand Hale (1983) sought to find the reason why UG appeared to allow any possible language to overtly determine grammatical functions in either a configurational or a non-configurational way, some years afterwards an attempt would be made not only to reduce the concept itself of grammatical function to more primitive terms but also, in the spirit of Chomsky (1981a), to eliminate all those particular rules which had been assumed to govern the grammatical function changing processes in favour of a P&P-based account. Such an ambitious task is undertaken in Mark Baker's book *Incorporation: a Theory of Grammatical Function Changing* (1988), a revised version of his PhD dissertation written in 1985. Similarly to Hale's (1983) paper, Baker's work makes use of a collection of empirical data from languages which are typologically distant from languages like English or French. However, here is where the similarity between these two works ends. In fact, Baker was not interested in comparing languages of the latter type with non-configurational ones, but rather with the configurational subtype going under the name of polysynthetic languages.

One of the most, if not the most, notable properties of polysynthetic languages is the occurrence of *incorporation processes* – namely, «processes by which one semantically independent word comes to be “inside” another» (Baker 1988, p. 1). As stated by Baker, the possibility of generating such linguistic constructions is not determined by a construction-specific rule, but rather is «no more than the result of applying standard movement transformations to words rather than to full phrases» (*ibidem*). In more technical terms, incorporation therefore consists in the application of the rule Move  $\alpha$  to a  $X^0$  element which, after leaving a trace in its base position, is adjoined to a higher  $Y^0$  element and, as a result, occupies a syntactic position which turns out to be the sister node of the latter. In the third chapter of his book, Baker deals specifically with what he calls “noun incorporation”, an example of which can be found in the following sentence from Onondaga, an American Indian language (Baker 1988, p. 76):

- (177) Pet wa?-ha-**hwist**-ahtu-?-t-a?. (Woodbury 1975)  
Pat PAST-3mS-**money**-lost-CAUS-ASP  
“Pat lost money.”

As far as (177) is concerned, the effect of noun incorporation can be seen from the fact that while the subject consists of one independent lexical item, the verb appears as a complex predicate containing both the verb root *ahtu* “lose” and the noun root *hwist* “money” all combined into one complex morphological unit. This kind of structure sharply contrast with sentence (178), which hosts a subject, a verb and a direct object all realized independently from one another (Baker 1988, p. 77):

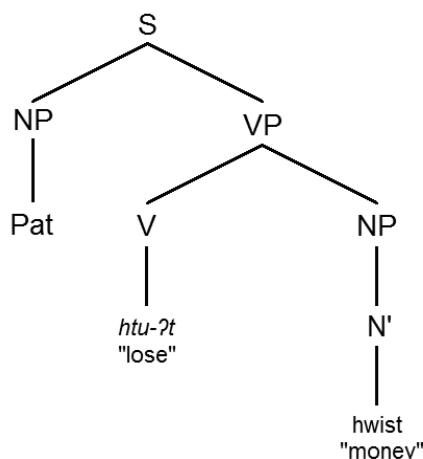
- |       |                               |     |                        |
|-------|-------------------------------|-----|------------------------|
| (178) | Pet waʔ-ha-htu-ʔt-aʔ          | neʔ | o- <b>hwist</b> -aʔ.   |
|       | Pat PAST-3mS/3n-lost-CAUS-ASP | the | PRE- <b>money</b> -SUF |
|       | “Pat lost the money.”         |     |                        |

After noting that in (177) «the noun root seems to count as the direct object of the structure, productively receiving a thematic role from the verb root» exactly as in (178) (Baker 1988, p. 76), Baker argues that such pairs of sentences can effectively be regarded as “thematic paraphrases” of one another, as the same theta-roles are being assigned from the same verb to the same arguments (cf. Baker 1988, pp. 77-78). This is indeed a plausible statement, fitting well into Chomsky’s (1981a) assumption that each argument is assigned a specific theta-role according the structural position it occupies at the level of D-structure. However, Baker further hypothesises that an argument bearing a specific theta-role always occurs in the same D-structure position with respect to the verbal head, thus implying a one-to-one relationship between each individual verb and the argument structure it projects. This hypothesis, which is crucial for Baker’s analysis of incorporation structures, is formulated as the Uniformity of Theta Assignment Hypothesis:

- (179) The Uniformity of Theta Assignment Hypothesis (UTAH):  
 Identical thematic relationships between items are represented by identical structural relationships between those items at the level of D-structure.  
 (Baker 1988, p. 46)

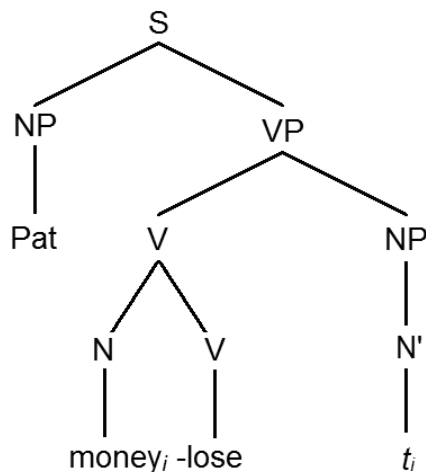
According to the UTAH, all cases of thematic paraphrases, including those represented by an incorporated structure and its non-incorporated counterpart, do actually share the very same D-structure. Turning to the case at hand, this implies that both (177) and (178) have the D-structure (180), with the subject “Pat” and the object “money” being respectively generated as the external and the internal arguments of the verbal head (Baker 1988, p. 80):

(180)



Assuming that noun-incorporation corresponds to the movement of the N element into V, if on the one hand the structure in (178) does not require any subsequent syntactic change with respect to (180), «in [177], however, the verb “lose” and the noun root “money” combine into a single word at some stage [...] by Move Alpha, which moves the structurally lower lexical item (the noun) to adjoin it to the higher lexical item in the syntax» (Baker 1988, p. 80). Consequently, according to Baker the S-structure of (177) is (181) (*ibidem*):

(181)



Another important element in Baker (1988) is the fact that incorporation, being an instantiation of head movement, obeys the same universal constraints applying to syntactic movement in general, such as the ECP. Considering the incorporation phenomena observed so far, this can be easily seen by noting that, as pointed out by Baker, «in ordinary transitive clauses, the direct object may be incorporated, but the subject may not be» (Baker 1988, p. 81), as shown in the following example taken from Mohawk (Baker 1988, pp. 81-82):

- (182) a. Yao-wir-a? a      ye-nuhwe? -s      ne ka-nuhs-a?. (Postal 1962)  
              PRE-baby-SUF 3fS/3n-like-ASP the PRE-house-SUF  
              “The baby likes the house.”
- b. Yao-wir-a? a      ye-**nuhs**-nuhwe? -s.  
              PRE-baby-SUF 3fS/3n-**house**-like-ASP  
              “The baby house-likes.”
- c. \*Ye-**wir**-nuhwe? -s ne ka-nuhs-a?.  
              3fS/3n-**baby**-like PRE-house-SUF  
              “Baby-likes the house.”

The nature of this subject-object asymmetry in noun incorporation can in fact be accounted for by assuming that, given for (182a) a D-structure representation analogous to (180), only the arguments which are generated under VP as direct objects of transitive verbs (but also as subjects of unaccusative verbs) will be able to c-command – and thus to antecedent govern – their trace after moving upward to the V node. On the other hand, any argument generated VP-externally (as in this particular case, the subject of a transitive verb) would have to move downward in the phrase marker in order to incorporate into V. Hence the impossibility for its trace to be properly governed and the consequent violation of the ECP in case of subject incorporation (cf. Baker 1988, p. 83).

In line with this latter view, another consequence of Baker’s approach is that incorporation is sensitive to the presence, between the incorporated  $X^0$  element and its trace, of any maximal projection acting as a potential barrier to antecedent-government, so that «there must be no BARRIER category that intervenes between the two» (Baker 1988, p. 55). Although already being introduced by Chomsky (1986a), in Baker (1988) the notion of barrier is implemented in a more flexible form in order to explain all GF changing phenomena in terms of incorporation, thus ideally reducing them to instances of head movement. More precisely, if on the one hand «for Chomsky, barriers are relative only to the potentially governed element», according to Baker «they be doubly relativized with respect to both the potential governor and the potential governee in the following way» (Baker 1988, p. 56):

- (183) Let D be the smallest maximal projection containing A. Then C is a BARRIER between A and B if and only if C is a maximal projection that contains B and excludes A, and either:
- (I)      C is not selected, or

- (II) the head of C is distinct from the head of D and selects some WP equal to or containing B.  
(Baker 1988, p. 56)

In these terms, barrierhood effects can arise in two distinct scenarios. According to (183i), government is blocked in those structures in which the XP headed by the incorporated  $X^0$  base position is not selected by the incorporating  $Y^0$  node, that is, whenever the mentioned XP is an adjunct. In the case of noun incorporation, the fact that «NI should never be able to take a noun root out of an NP adjunct that appears in the VP» can be better understood resorting to the minimalist account of clause representation, as according to this model the XP adjunct and the VP it adjoins to are represented as sister nodes, thus excluding any c-command relation between the incorporating head V and (any of the nodes in) the adjoined phrase (Baker 1988, p. 86). Furthermore, (183ii) states that «an intervening theta assigner also breaks a government path» between V and the incorporated element, exactly as Chomsky's (1986a) *Minimality Condition* holds of the following configuration (Baker 1988, p. 56):

- (184) ...  $\alpha$  ... [<sub>y</sub> ...  $\delta$  ...  $\beta$  ... ]  
(Chomsky 1986a, p. 42)

in which  $\alpha$  cannot govern  $\beta$  as the lexical head  $\delta$  yields a potential ambiguity effect in the government chain by being closer to  $\beta$  than  $\alpha$  is:

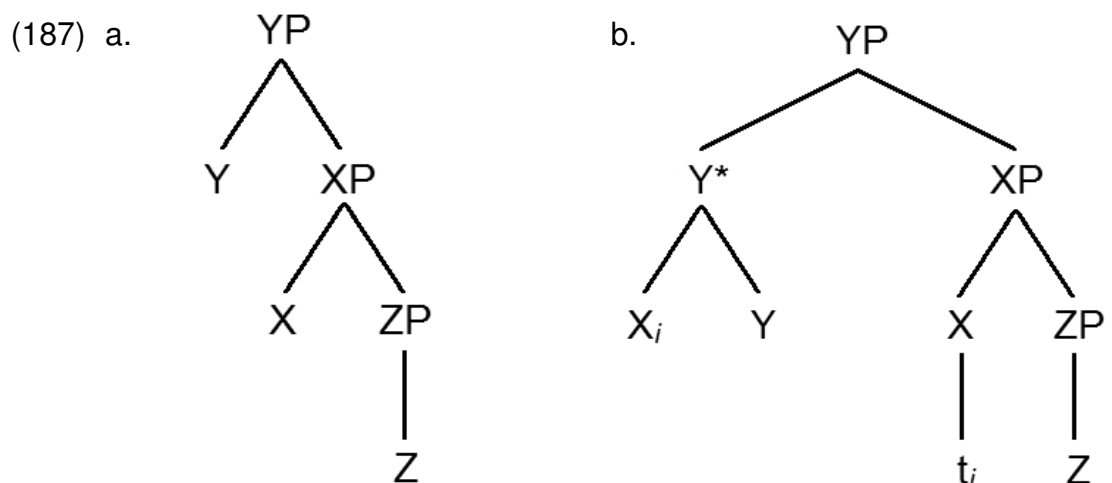
- (185) The Minimality Condition:  
y is a barrier for  $\beta$  if y is a projection or the immediate projection of  $\delta$ , a zero-level category distinct from  $\beta$ .  
(*ibidem*)

However, an important difference between Chomsky's Minimality Condition (185) and Baker's newly proposed formulation (183ii) lies in the way the notion of *distinctiveness* is respectively implemented by the two. In fact, while in the former it is the intervening element  $\delta$  which has to be distinct from  $\beta$  for Minimality to apply, in the latter this same requisite regards «the head of C» and «the head of D»; that is, referring now to (184) for the sake of clarity, the intervening element  $\delta$  and the potential governor  $\alpha$  (Baker 1988, p. 56). This difference is very important, as this seemingly small addition crucially allows Baker to link incorporation to GF changing processes and, at the same time, to find an

explanation to some otherwise unaccountable phenomena, like the one represented by the following example (Baker 1988, p. 65):

- (186) a. Ka-rakv ne [sawatis hrao-**nuhs**-a?]. (Postal 1962)  
 3n-white DET John 3m-house-SUF  
 b. Hrao-**nuhs** -rakv ne [sawatis t].  
 3m-house-white DET John  
 “John’s house is white.”

As noted by Baker, the unincorporated sentence (186a) has (187a) as its S-structure, while its incorporated counterpart (186b) has (187b), «where the verb *-rakv* “white” is “Y”, the noun *-nuhs*- “house” is “X”, and the NP *sawatis* “John” is “ZP» (Baker 1988, pp. 64-65):



Assuming with Baker that «a verb can only agree with an NP which it governs», the shift in verbal agreement between (186a) and (186b) is rather puzzling (Baker 1988, p. 65). More precisely, the fact that in (186a) the verb *-rakv* has neuter agreement matching “house”, whereas in (186b) it has masculine agreement matching “John”, seems compatible with a change in the government relations between the verb and its arguments. Furthermore, such a shift also seems to run counter to Chomsky’s Minimality Condition, as the head X intervening between Y and Z should give rise to a barrier within its maximal projection for its complement phrase which, in turn, should block government between the verb and “John” in (186b). However, according to Baker’s (183ii), the derivation of (186b) can be directly accounted for by assuming that when N undergoes incorporation into V, the

movement chain which is established between the trace  $t_i$  and its antecedent  $X_i$  within the complex verb makes Y and the head of the potential barrier XP no more distinct from each other. In these terms, Baker's notion of distinctiveness corresponds to the concept of co-indexing independence, which is applied to X and Y in the following way (Baker 1988, p. 64):

- (188) X is **distinct** from Y only if no part of Y is a member of a (movement) chain containing X.

As a consequence, according to Baker the agreement shift between (186a) and (186b) does indeed follow from a difference in the government properties of their respective grammatical structures: on the one hand, in (186a) the verb cannot agree with the noun "John" because the two are separated by a maximal projection whose head, "house", is distinct from the potential governor and hence acts as an intervening barrier; on the other hand, in (186b) the head of the potential barrier is incorporated into the verb, with the result that the former is no more distinct from the latter and the verb governs "John", thus allowing agreement between the two (cf. Baker 1988, p. 65). Far from being regarded as a case-particular solution, the fact that after incorporation the complex verb governs every element the incorporated head governed at D-structure is systematized by Baker into one principle which is essential not only for allowing the possibility of a set of incorporation phenomena which, like (186a-186b), would otherwise be impossible in terms of the Minimality Condition, but also for accounting for the inner relationship between GF changing phenomena and incorporation processes. This principle is formulated as the Government Transparency Corollary:

- (189) The Government Transparency Corollary (GTC):  
A lexical category which has an item incorporated into it governs everything which the incorporated item governed in its original structural position.  
(Baker 1988, p. 64)

The central role of the GTC in Baker (1988) is also exemplified in the analysis of causative constructions, which takes place in the fourth chapter. According to Baker, in polysynthetic languages «morphological causatives are [...] exactly like Noun Incorporation, except for the category of the word being moved», which in this case corresponds to a verb rather than a noun (Baker 1988, p. 149). Consequently, these



constructions are nothing more than instances of verb incorporation and can therefore be accounted for by the same independent principles of grammar which also govern noun incorporation phenomena:

In these constructions, a single verb corresponds not to a verb and a noun, but rather to two verbs. This possibility, together with Noun Incorporation, is the second major element of polysynthesis. Here again, we will find strong evidence that the forms are actually syntactically derived from two independent verbs by movement. Thus, causatives are VERB INCORPORATION (VI), directly parallel to Noun Incorporation and subject to exactly the same principles. (Baker 1988, p. 147)

In polysynthetic languages such as the Bantu language Chichewa, the fact that causative constructions can be realized as either unincorporated or incorporated structures does constitute, on par with the cases of optional noun incorporation (177-178), an example of thematic paraphrase. This can be seen in the following pair of sentences (Baker 1988, p. 148):

- (190) a. Mtsikana ana-chit-its-a            kuti mtsuko    u-gw-e. (Trithart 1977)  
           girl            AGR-do-make-ASP that waterpot AGR-fall-ASP  
       b. Mitsikana anau-gw-ets-a            mtsuko.  
           girl            AGR-**fall-made**-ASP waterpot  
           “The girl made the waterpot fall.”

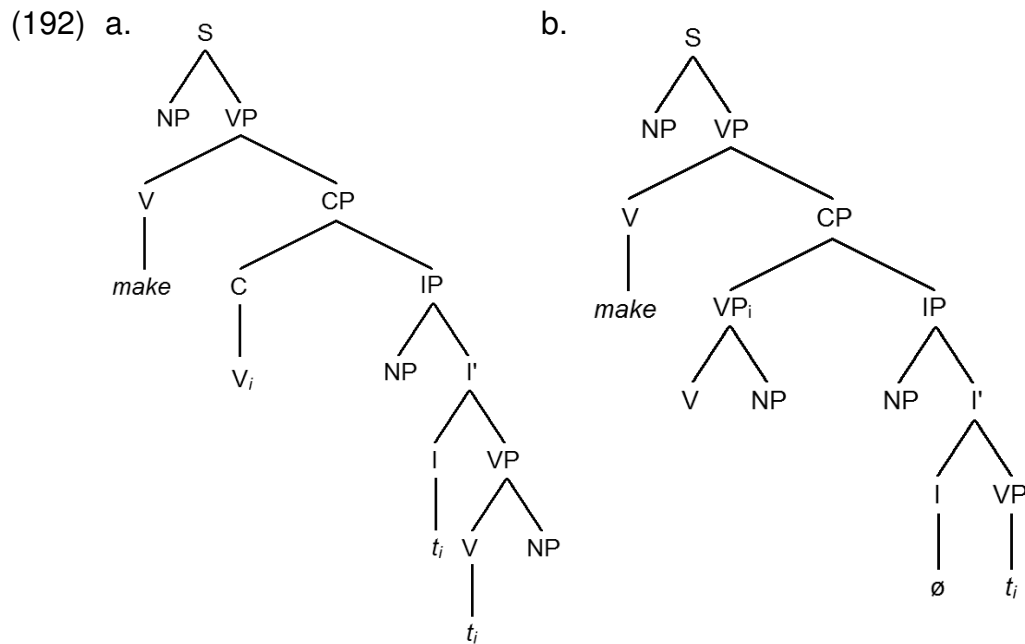
Baker’s basic intuition concerning the status of morphological causatives is that, along with the fact that (190a) and (190b) share the same D-structure, the Chichewa causative morpheme *-its* «is an affix, and hence has a morphological subcategorization frame which stipulates that it must attach to a verb [...] before S-structure» (Baker 1988, p. 151). In order to assure this, the causative verb can trigger either the insertion of a pleonastic *do* acting as its verbal root or the incorporation of a lexical verb, which is moved from the embedded S into the matrix sentence’s causative verb via  $X^0$  movement. These two options are shown in (191a) and (191b) respectively (*ibidem*):

- (191) a. [<sub>S</sub> girl Infl do+*its* [<sub>S</sub> waterpot Infl fall]]  
       b. [<sub>S</sub> girl Infl fall<sub>i</sub>+*its* [<sub>S</sub> waterpot Infl *t<sub>i</sub>*]]

Another aspect which emerges as crucial in Baker's analysis of morphological causatives is that noun incorporation and verb incorporation structures differ in their respective complexity. More precisely, if on the one hand noun incorporation relates a complex verb and an  $X^0$  element both occurring within the same minimal clause, by assuming the correctness of (191b) verb incorporation involves  $X^0$  movement of a complement verb generated in the embedded clause into a causative verb belonging to the matrix clause. In these terms, in the latter case the complex verb and the incorporated  $X^0$  element are separated by a CP and an IP which, by acting as two potential barriers to government, do not allow the complement verb to reach the matrix verb by means of a single  $X^0$  movement:

Suppose that causative morphemes are like other elements that take propositional complements in that they subcategorize for a full S'. Then, the matrix verb does not govern the embedded verb, because the maximal projections of C and I intervene, both of which are barriers because their heads select a phrase which contains the lower verb (IP and VP respectively) Thus, if the embedded verb is moved directly onto the matrix verb, it will not govern its trace, and the structure will be ruled out by the ECP. (Baker 1988, p. 168)

Baker argues that morphological causative constructions overcome this limitation in two ways. According to the first solution, the embedded verb «reaches the C position by incorporating first into the embedded I» (Baker 1988, p. 170). In this way, the head of IP will be no more distinct from the head of CP and IP will therefore not act as a barrier between C and the lower verb. After incorporating into I and C, the lower V can thus move to the higher V. According to the second solution, assuming that «the specifier of C' [...] is a maximal projection position by X-bar theory», the embedded verb moves to this position by taking its entire VP projection along with it (*ibidem*). While the VP constitutes a potential barrier between the incorporating verb and the verb to be incorporated, it does ultimately not act as such as «its head is not distinct from the antecedent or the trace, and it is not an adjunct» (Baker 1988, p. 171). Therefore, after moving into the specifier of C, the embedded verb can finally move into the matrix verb. These two possible intermediate structures underlying verb incorporation are shown here (Baker 1988, p. 173):



Although the parallel possibility of these two movement sequences seems to represent an unnecessary complication, Baker attributes it to a case of cross-linguistic variation regarding the existence of two distinct types of causative constructions:

Morphological causative constructions, although biclausal semantically and underlyingly, appear monoclausal on surface. Causative constructions then vary as to which of the NPs from the embedded clause acts like the direct object in this single surface clause (Baker 1988, p. 162)

The first causative construction, which corresponds to languages such as standard Chichewa (Chichewa-A), the embedded “ergative” (here meant as the transitive subject) surfaces as an oblique, indirect object, while the embedded “absolutive” (a term used by Baker to refer to the transitive object and the intransitive subject) surfaces as the direct object. This type is exemplified by the following sentences in Mchombo, which feature an embedded intransitive verb and an embedded transitive verb respectively (Baker 1988, pp. 162-163):

- (193) a. Boma            li-ku-sow-ets-a                                  nsomba.  
                  government SP-PRES-disappear-CAUS-ASP fish  
                  “The government made fish disappear (become unavailable).”
- b. Anyani    a-na-meny-ets-a                                  ana    kwa buluzi.  
                  Baboons SP-PAST-hit-CAUS-ASP children to    lizard  
                  “The baboons made the lizard hit the children.”

In the second causative type, which is displayed by the Bantu dialect which Baker calls Chichewa-B, the embedded subject – whether «used transitively or intransitively» – becomes the direct object of the causative verb while the embedded object, if present, «surfaces as a kind of second object», as in the following example (Baker 1988, p. 164):

(194) Catherine a-na-kolol-ets-a mwana wake chimanga.  
Catherine SP-PAST-harvest-CAUS-ASP child her corn  
‘Catherine made her child harvest the corn.’ (Trithart 1977)

Considering these linguistic facts, Baker's hypothesis is that the two intermediate structures represented in (192) correspond to the two attested patterns of argument realization in morphological causatives. Moreover, at a deeper level, this optionality can be regarded as the result of the interaction between the universal principles of Case Theory and «the existence of marked types of Case assignment», in turn concerning complex verbs derived by verb incorporation processes, whereby languages such as Chichewa-A and Chichewa-B differ idiosyncratically (Baker 1988, p. 173).

Differently from Chichewa-A, which represents unmarked Case assignment as its verbs assign only one accusative Case, languages of the same type as Chichewa-B «appear to be marked in that (some of) their verbs can assign structural Case to more than one NP which they govern» (Baker 1988, p. 174). Starting with the unmarked case, according to Baker the only derivation possible for morphological causatives in non-double object languages such as Chichewa-A is the one corresponding to (192b), and this because this intermediate structure is the only one which allows both the satisfaction of the requirement for what he calls «adjacency between Case-indexed items at PF» and the Case Filter to not be violated (Baker 1988, p. 173). The former requisite, which crucially regards the S-structure position of the embedded “absolutive” NP with respect to the complex verb which assigns its only structural Case to it, is in fact satisfied by moving the lower V to SPEC-C through XP movement of its entire maximal projection. On the other hand, the Case Filter violation which would be triggered by the Case-less embedded “ergative” NP is avoided by assigning oblique Case to the latter through the application of a rule of preposition insertion (cf. Baker 1988, pp. 188-189).

Concerning the “double accusative” verbs featured in Chichewa-B and assuming that «directed strict adjacency will not be a requirement for the PF interpretation of Case assignment for at least one of the structural Cases in such a language, since both cannot

be adjacent to the verb», the embedded object cannot be liable for a violation of this condition whatever its position in the embedded S' (Baker 1988, p. 174). Moreover, since after incorporation the complex verb governs all the nodes in the embedded S' as a consequence of the GTC, the successive cyclic movement of the lower verb into I, C and then into the matrix V implies that both the embedded NP arguments will receive the two available structural accusative Cases. Hence the possibility of (192a) and, therefore, the need of having two parallel intermediate structures for morphological causatives.

At this point, Baker delves into the analysis of causative constructions in Romance languages. As stated by Baker, a common preliminary observation is that Romance causatives behave exactly as Chichewa-A's morphological causatives in their argument realization patterns. For example, this is shown in Italian (Baker 1988, p. 201):

- (195) a. Maria fa          lavorare Giovanni. (Burzio 1986)  
           Maria makes work      Giovanni  
           “Maria makes Giovanni work.”  
       b. Maria ha fatto riparare la macchina a Giovanni.  
           Maria has made fix      the car          to Giovanni  
           “Maria made Giovanni fix the car.”

Being the subject of an intransitive verb, the embedded subject *Giovanni* in (195a) is realized as an accusative direct object. On the other hand, since in (195b) the lower verb is transitive, the embedded subject *Giovanni* is realized as an oblique object. Therefore, the argument realization of Italian causatives exactly follows the pattern shown in (193). Having said that, examples like (195) show that, despite their striking similarities with non-double object polysynthetic languages, Romance languages do not allow the same kind of verb incorporation. In fact, in Italian «the causative verb *fare* and the lower verb simply do not become a single word morphologically» (Baker 1988, p. 202):

This collection of facts suggests that we must give an account of Romance causatives in which they have exactly the same syntax as (say) Chichewa causatives, but they differ with respect to the morphology. (*ibidem*)

Generalizing to Romance the correlations between Case marking and causative construction type which applies to Chichewa-A, Baker's intuition is that examples like (195) «seem to be cases of “incorporation” without the incorporation» (*ibidem*). In these terms,

Baker's idea is that the very same incorporation process which combines a verb and a structurally lower lexical element can happen either in the mapping between D-structure and S-structure or between S-structure and LF. In the first case, which corresponds to what happens in polysynthetic languages, the verb and its incorporated object form a single morphological word which is overtly realized as a complex verb. In the second case, which Baker calls *abstract incorporation* (or *reanalysis*, explicitly referring to the approach followed by Hornstein & Weinberg (1981) and which has briefly been presented above in connection with Kayne (1983)), although no actual combination of morphological forms is visible «the coindexing between the nodes is interpreted exactly like the coindexing relationship between a complex word and the trace of one of its parts» (Baker 1988, p. 202). The difference between incorporation and reanalysis is represented in the following schemata (*ibidem*):

- (196) a.  $[_{YP} \dots [X_i + Y]_Y \dots [_{XP} t_i \dots]]$   
 b.  $[_{YP} \dots Y_i \dots [_{XP} X_i \dots]]$

It is in this way that Baker accounts for the fact that Italian causatives and Chichewa-A causatives are syntactically identical. Interestingly, Baker interprets the requirement for reanalysis in Italian causative constructions to what he calls «a semi-semantic property of the [causative] verb» by indicating the causative verb *fare* as a “reanalyzer” rather than an “incorporator” (Baker 1988, p. 203), and hence attributing the optionality between regular incorporation and abstract incorporation in causative constructions to a lexically-based idiosyncratic property similar to the one argued in Chomsky (1981a) with respect to ECM constructions.

One of the most, if not the most, crucial aspect of Baker's distinction between regular incorporation and abstract incorporation is that this duality plays a central role not only in verb incorporation, but also in all the other possible instances of incorporation regardless of the moved  $X^0$  element's syntactic category. Leaving aside considerations regarding noun incorporation phenomena, which would be too lengthy to discuss here, some specific cases of preposition incorporation can be useful to better understand the theoretical scope of Baker's analysis within the P&P model. In this regard, in the fifth chapter of Baker (1988) some attention is given to the fact that, along with cases of regular preposition incorporation, «there also exist instances of Preposition Reanalysis [...] having the same relation to the former as Italian causatives have to Chichewa or Malayalam

causatives» (Baker 1988, p. 259). In this regard, one of the examples of Preposition Reanalysis which are analyzed by Baker is the pseudo-passive construction (*ibidem*):

- (197) a. Everyone talked about Fred.  
b. Fred was talked about (last night).

As can easily be seen in (197), in the pseudo-passive construction «the NP which was the object of a preposition becomes the subject when the main verb of its clause is put into the passive» (*ibidem*). While in English pseudo-passives are possible, in most languages they are not. This is specifically true for French, as shown below (*ibidem*):

- (198) a. Tout le monde a parlé de Fred. (Kayne 1983)  
b. \*Fred a été parlé de (hier soir).

In order to account for this cross-linguistic difference, Baker attributes the impossibility of forming pseudo-passives to the unavailability, in languages like French, of a reanalysis procedure whereby the preposition is abstractly incorporated into the verb at LF. According to Baker, the ungrammaticality of (198b) is in fact due to the fact that in (198) the NP object of the preposition is not governed by V «since government is blocked by the P» (Baker 1988, p. 260). Therefore, in (198a) movement of the NP *Fred* to subject position would induce an ECP violation. However, in line with what is argued in Kayne (1983), in (197b) P undergoes abstract incorporation with V, with the consequence that as predicted by the GTC, «the verb complex governs what the P governed before it moved», including the trace left by the NP object (*ibidem*):

In other words, the English constructions have the properties of Preposition Incorporation, but without the actual incorporation – which is exactly the characterization of the Reanalysis relation. (*ibidem*)

Although Kayne (1983) does not attribute the availability of V-P reanalysis to the effect of a syntactic parameter, it can be argued that Baker (1988) does. Throughout Baker (1988) special emphasis is put on the fact that some relevant language-specific properties interacting with UG's subcomponents like Case Theory do actually correspond to differently settable parameters. One notable example is the marked type of Case assignment distinguishing Chichewa-B from Chichewa-A which, although never explicitly

defined as a parametric option, is discussed in a paragraph titled “Case Parameters and Causative Variation” (cf. Baker 1988, p. 161). Regarding the occurrence of either syntactic incorporation or its LF counterpart, the parametric nature of this opposition is mentioned, although briefly, in connection with noun incorporation phenomena, as Baker expresses the conviction that «N-V Reanalysis does exist parallel to N Incorporation as an option in universal grammar» (Baker 1988, p. 274).

In conclusion, by attempting to reduce a rather comprehensive set of grammatical function changing phenomena to an instantiation of the basic syntactic operation Move  $\alpha$ , Baker (1988) contributed to the development of the P&P model also by hypothesizing the existence of a binary difference in the way incorporation, when allowed, can change the relationship between the verb and its pattern of government.

### 2.5.3 – *The parametrization of the notion of governing category*

Although the formulation of the binding principles constituted one of the main achievements in Generative Grammar, by the mid-Eighties a number of binding phenomena had not been accounted for by the theory set forth in Chomsky (1981a). In this respect, in their paper *Parameters, binding and learnability* (1987) Maria Rita Manzini and Kenneth Wexler sought to regularize some of the attested exceptions to Principles A and B of Binding Theory by proposing the existence of two parameters of variation which would interact with both the notion of governing category and antecedency requirements.

In the very first part of their paper, Manzini & Wexler introduce an approximation of the definition of governing category as originally proposed in Chomsky (1981a):

- (199)  $\gamma$  is a governing category for  $\alpha$  iff  
 $\gamma$  is the minimal category that contains  $\alpha$  and a governor for  $\alpha$  and has a subject.  
 (Manzini & Wexler 1987, p. 416)

where the term “subject” in (199) stands for either the NP subject of an infinitive or the INFL element of a tensed sentence (cf. Chomsky 1993 [1981a], pp. 209-210). Together with Principles A and B of Binding Theory, the definition given in (199) is applied to the following examples from Italian, which feature the reflexive *sé* (Manzini & Wexler 1987, p. 416):



- (200) *Alice<sub>j</sub> sapeva [<sub>V</sub> che Mario<sub>i</sub> aveva guardato sé<sub>i/\*j</sub> nello specchio].*  
 Alice knew that Mario had looked at REFL in the mirror
- (201) *Alice<sub>j</sub> pensava [<sub>V</sub> che Mario<sub>i</sub> avesse guardato sé<sub>i/\*j</sub> nello specchio].*  
 Alice thought that Mario had (subj.) looked at REFL in the mirror
- (202) *Alice<sub>j</sub> vide [<sub>V</sub> Mario<sub>i</sub> guardare sé<sub>ij</sub> nello specchio].*  
 Alice saw Mario look at REFL in the mirror
- (203) *Alice<sub>j</sub> guardò [<sub>V</sub> i ritratti di sé<sub>ij</sub> di Mario<sub>i</sub>].*  
 Alice looked at portraits of REFL of Mario  
 “Alice looked at Mario’s portraits of REFL.”

According to (199), in (200-203) the reflexive’s governing category corresponds to the embedded sentence. This does not constitute a problem for Principle A, as in these examples the reflexive *sé* is only bound by *Mario*, that is, its only possible antecedent within its governing category. However, since in (202) and (203) the minimal XPs containing the anaphor, its governor and a subject are the embedded clause and the embedded nominal respectively, *sé* has a potential antecedent also outside its governing category, hence contradicting Principle A. At this point it is worth observing that the embedded clause in (202) and the embedded nominal in (203) differ from (200) and (201) in lacking an INFL node. Consequently, if having an INFL node were an additional syntactic requirement in the definition of governing category as formulated in (199), in both (202-203) the minimal XP containing the reflexive and the governor for the reflexive would be the matrix sentence, with *Alice<sub>j</sub>* thus being able to be an antecedent for *sé* without giving rise to ungrammaticality. This is precisely the provisional solution adopted in Manzini & Wexler (1987), who accordingly reformulated the definition of governing category as follows:

- (204)  $\gamma$  is a governing category for  $\alpha$  iff  
 $\gamma$  is the minimal category that contains  $\alpha$  and the governor for  $\alpha$  and has an  
 Infl. (Manzini & Wexler 1987, p. 416)

As can be seen in (204), the requirement “has a subject” of (199) can simply be replaced by the requirement “has an Infl”, as all categories featuring an INFL node correspond to sentences. This same reasoning is also applied to all the other governing categories which are postulated throughout Manzini & Wexler (1987), as such categories are specifically assumed not only to entertain, but also to be «ordered with respect to one

another by the set-theoretical relation of proper inclusion» (Manzini & Wexler 1987, p. 419).

The subsequent step of Manzini and Wexler argumentation consists, first, in analyzing further sets of empirical cases, each corresponding to a complex sentence featuring either an anaphor or a pronominal taking at least two potential antecedents, one inside and the other outside their proposed governing category. Then, a different parametrization of the notion of governing category is proposed on a case-by-case basis in such a way as to prevent any violation of the Binding Theory. After Italian reflexive *sé*, Manzini and Wexler analyze the behaviour of the Icelandic reflexive *sig* (Manzini & Wexler 1987, p. 417):

- (205) Jón<sub>j</sub> segir [<sub>v</sub> að María<sub>i</sub> elskar sig<sub>//\*</sub>].  
 John says that Maria loves REFL
- (206) Jón<sub>j</sub> segir [<sub>v</sub> að María<sub>i</sub> elski sig<sub>//j</sub>].  
 John says that Maria loves(subjunctive) REFL
- (207) [<sub>v</sub> María<sub>j</sub> skipaði Haraldi<sub>i</sub> að PRO raka sig<sub>//j</sub>].  
 Maria ordered Harald to shave REFL
- (208) [<sub>v</sub> Jón<sub>j</sub> heyrði lysingu Mariu<sub>i</sub> af ser<sub>//j</sub>].  
 John heard description Maria(gen) of REFL  
 “John heard Maria’s description of REFL.”

As pointed out by the authors, «neither the definition of governing category in [199] nor the definition of governing category in [204] gives the correct results with respect to these data» (*ibidem*). For example, if on the one hand (204) does not give rise to any problem with respect to (205), (207) and (208), on the other hand in (206) the minimal category that contains *sig* and a governor for *sig* and has an INFL is the embedded sentence. However, contrary to Principle A, the co-indexing between *sig* and the matrix subject in (206) does not give rise to ungrammaticality. Similarly as done for (200-203), the descriptive value of Principle A can be preserved for all of (205-208) by postulating a further definition of governing category expressly built on the requirement of having a “referential” Tense – that is, an indicative Tense «whose properties are inherently defined» as opposed to a subjunctive Tense, whose properties in turn depend anaphorically upon a superordinate Tense (*ibidem*):

- (209)  $\gamma$  is a governing category for  $\alpha$  iff  
 $\gamma$  is the minimal category that contains  $\alpha$  and a governor for  $\alpha$  and has a  
 “referential” Tense.  
 (Manzini & Wexler 1987, p. 417)

Under the definition in (209), in (206) the governing category of *sig* extends to the matrix sentence, thus including both potential antecedents *Maria* and *Jon*.

As anticipated above, the same approach adopted so far for the analysis of anaphors is also implemented for evaluating the behaviour of pronominals with respect to Principle B. In this sense, the only pronominal considered in Manzini & Wexler (1987) is the Icelandic *hann* (Manzini & Wexler 1987, p. 418):

- (210) Jón<sub>i</sub> segir að María elskar hann<sub>i</sub>.  
 John says that Maria loves him  
 (211) Jón<sub>i</sub> segir að María elski hann<sub>i</sub>.  
 John says that Maria loves(subjunctive) him  
 (212) \*Jón<sub>i</sub> skipaði mér að raka hann<sub>i</sub>.  
 John ordered me to shave him

As noted by the authors, «by the definitions of governing category in [199] and [204], the governing category for *hann* is the embedded sentence in [210]-[212]» (*ibidem*). As a consequence, while in (210-211) Principle B is fulfilled, the same principle is violated in (212) as *hann* cannot be bound by the matrix subject *Jón*. A similar problem arises if adopting the definition of governing category in (209). More precisely, since the minimal XP containing *hann*, the governor for *hann* and a “referential” Tense would be the embedded sentence in (210) and the matrix sentence in (211-212), although (210) and (212) would not represent a problem on the other hand the grammaticality of (211) would imply a violation of Principle B due to the impossibility for the pronominal *hann* to be free in its governing category. A new definition of governing category is therefore considered:

- (213)  $\gamma$  is a governing category for  $\alpha$  iff  
 $\gamma$  is the minimal category that contains  $\alpha$  and the governor for  $\alpha$  and has a  
 Tense.  
 (*ibidem*)

The definition in (213) correctly accounts for all of (210-212). As far as (210-211) are concerned, the governing category of *hann* is the embedded sentence, with the antecedent *Jón* being in a proper position outside of it. Regarding (212), here the minimal category containing a pronominal and a governor for it and having a Tense is the matrix sentence, hence the violation of Principle B ruling out the sentence.

The last element considered for the formulation of Manzini & Wexler's (1987) governing category parameter is the Japanese reflexive *zibun* (Manzini & Wexler 1987, p. 419):

(214) John-wa<sub>j</sub> [Bill-ga<sub>i</sub> zibun-o<sub>ij</sub> nikunde iru] to omotte iru.

John Bill Refl hates that thinks

"John thinks that Bill hates Refl."

(215) John-wa<sub>j</sub> [Bill-ga<sub>i</sub> zibun-o<sub>ij</sub> syasin-o mihatte iru] to omotte iru.

John Bill Refl pictures is watching that thinks

"John thinks that Bill is watching pictures of Refl."

While «none of the definitions of governing category given so far correctly accounts for these data», according to Manzini & Wexler the definition in (216) correctly predicts both (214) and (215) (*ibidem*):

(216)  $\gamma$  is a governing category for  $\alpha$  iff

$\gamma$  is the minimal category that contains  $\alpha$  and the governor for  $\alpha$  and has a "root" Tense.

(Manzini & Wexler 1987, p. 418)

Since a governing category whose characterizing property is to have a "root" Tense does necessarily correspond to the matrix sentence, *zibun* is predicted to be potentially co-indexable by any R-expression c-commanding it. Hence the possibility that either the matrix subject or the embedded object may bind *zibun* in (214-215).

At this point, the five different definitions of governing category proposed so far are provisionally combined into a syntactic parameter featuring five different values:

- (217)  $\gamma$  is a governing category for  $\alpha$  iff  
 $\gamma$  is the minimal category that contains  $\alpha$  and a governor for  $\alpha$  and has  
 a. a subject; or  
 b. an Infl; or  
 c. a Tense; or  
 d. a “referential” Tense; or  
 e. a “root” Tense.  
 (Manzini & Wexler 1987, p. 419)

Although (217) already accounts for the different behaviour of the various instances of anaphoric and pronominal binding shown above, in order to fully capture binding phenomena in all their possible range of diversity and complexity Manzini and Wexler assume that «some notion of accessibility is needed in the definition of governing category» (Manzini & Wexler 1987, p. 420). As pointed out by the authors themselves, the notion of accessibility dates back to Chomsky (1981a), who originally formulated it as an additional requirement for subjecthood in relation to the notion of governing category (here Chomsky’s original formulation is rewritten by using  $\gamma$  in place of  $\beta$  for clarity’s sake):

- (218)  $\gamma$  is a governing category for  $\alpha$  if and only if  $\gamma$  is the minimal category containing  $\alpha$ , a governor of  $\alpha$ , and a SUBJECT accessible to  $\alpha$ .  
 (Chomsky 1993 [1981a], p. 211)

As noted by Manzini and Wexler, «a category is accessible to an element  $\alpha$  iff it c-commands  $\alpha$  and it can be coindexed with  $\alpha$  without violating the *i*-within-*i* Condition», as shown in (219) (Manzini & Wexler 1987, p. 420). The *i*-within-*i* Condition, which was also formulated in Chomsky (1981a), in turn corresponds to the fact that «no element can be contained in a category bearing the same index», as shown in (220) (*ibidem*):

- (219)  $\beta$  [that is, the subject of a governing category] is accessible to  $\alpha$  iff  
 $\beta$  c-commands  $\alpha$  and  
 coindexing of  $\alpha$  with  $\beta$  does not violate the *i*-within-*i* Condition.  
 (220)  $*[{}_i \dots \alpha_i \dots]$   
 (*ibidem*)

While accepting (219-220), Manzini and Wexler not only impose a distinction between anaphors and pronominals whereby accessibility holds for the former elements but not for the latter but, differently from Chomsky (1981a), the accessibility requirement

invoked by them has to be met only by the subject of the anaphor's governing category – that is, without checking also other subjects (cf. Manzini & Wexler 1987, p. 420). In these terms, by combining the parametric definition of governing category in (217) with Chomsky's notion of accessibility the following definition is obtained:

- (221)  $\gamma$  is a governing category for  $\alpha$  iff
- $\gamma$  is the minimal category that contains  $\alpha$  and a governor for  $\alpha$  and
    - a. can have a subject or, for  $\alpha$  anaphoric, has a subject  $\beta$ ,  $\beta \neq \alpha$ ; or
    - b. has an Infl; or
    - c. has a Tense; or
    - d. has a "referential" Tense; or
    - e. has a "root" Tense;
  - if, for  $\alpha$  anaphoric, the subject of  $\gamma$  is accessible to  $\alpha$ .
- (Manzini & Wexler 1987, p. 421)

The advantage of assuming a definition of governing category incorporating accessibility with respect to (217) is evident when considering examples such as the following Italian sentence (Manzini & Wexler 1987, p. 422):

- (222) Alice<sub>i</sub> sapeva che [[i miei ritratti di sé<sub>ij</sub>] spaventavano Mario<sub>j</sub>].  
 Alice knew that my portraits of REFL frightened Mario

As already pointed out for (200-203), the reflexive *sé* requires its governing category to be specified as having an INFL node. Consequently, in principle *sé* would now abide by either clause (b) of (217) (which in turn corresponds to definition (204)) or, if accessibility actually plays a role in Binding Theory, clause (b) of (221). According to the former definition, as noted by Manzini and Wexler the governing category for *sé* in (222) is the embedded sentence. This yields an undesirable result, as *sé* can be bound by either the matrix subject *Alice*, which occurs outside the embedded sentence, or the embedded object *Mario* which, crucially, does not c-command the anaphor it may be co-indexed with. On the other hand, according to the definition in (221) *sé* does not have a governing category at all, the reason being that «the subject of the embedded sentence» *i miei ritratti di sé* «is not accessible to *sé*» (*ibidem*). Hence, having no governing category, «*sé* is predicted not to be subject to any binding condition», thus being able to share its referential index with both extra-sentential and non-c-commanding antecedents (*ibidem*).

Despite the fact that (221) accounts for examples such as (222), there is still another clause needed by what in Manzini and Wexler (1987) is assumed to be the right definition of governing category. Attention is solicited to the following example from Korean (Manzini & Wexler 1987, p. 422):

- (223) John-in<sub>k</sub> [[Bill-i<sub>i</sub> caki-lil<sub>ij/k</sub> miwəha-nin-kəs-i] Mary-lil<sub>j</sub> silphike ha-əss-ta] mit-nin-ta.  
 John Bill Refl hates Mary sad made thinks  
 “John thinks that it made Mary sad that Bill hates Refl.”

By applying (221) under clause (e), in the above example the governing category for the reflexive *caki* corresponds to the root sentence. As the matrix subject *John* is accessible to *caki*, as pointed out by Manzini & Wexler «*caki* is predicted to be bound in the root sentence [...] incorrectly, since *caki* can have an antecedent that does not c-command it, as in the case with the embedded object» (*ibidem*). However, by modifying the accessibility clause in (221) with the requirement that «not only the subject of  $\gamma$  but also the subject of any category intervening between  $\alpha$  and  $\gamma$  must be accessible to  $\alpha$ », in (224) *caki* would be predicted to not have any governing category, hence being able to pick up its antecedent freely (*ibidem*). In these terms, Manzini & Wexler’s final definition of governing category is proposed:

- (224)  $\gamma$  is a governing category for  $\alpha$  iff  
 $\gamma$  is the minimal category that contains  $\alpha$  and a governor for  $\alpha$  and  
 a. can have a subject or, for  $\alpha$  anaphoric, has a subject  $\beta$ ,  $\beta \neq \alpha$ ; or  
 b. has an Infl; or  
 c. has a Tense; or  
 d. has a “referential” Tense; or  
 e. has a “root” Tense;  
 if, for  $\alpha$  anaphoric, the subject  $\beta'$ ,  $\beta' \neq \alpha$ , of  $\gamma$ , and of every category dominating  $\alpha$  and not  $\gamma$ , is accessible to  $\alpha$ .  
 (Manzini & Wexler 1987, pp. 422-423)

Under (224), in (223) the subject of the intermediate embedded sentence is not accessible to *caki*, as it occurs in a non-c-commanding position. As a result, *caki* is able to be co-referentially dependent upon any potential antecedent, thus vacuously satisfying Principle A.

In addition to (224), which is explicitly referred to as the *governing category parameter* (cf. Manzini & Wexler 1987, p. 429), in order to fully capture the distribution of both anaphors and pronominals in Manzini & Wexler (1987) another syntactic parameter is proposed. In particular, the formulation of Manzini & Wexler's (1987) second parameter arises from a striking difference which can be observed, for example, between English and other languages like Japanese and Icelandic and which concerns both anaphors and pronominals. First, although «the English anaphors *himself* and *each other* can be bound in their governing category by subjects and nonsubjects alike», Japanese *zibun* must be bound in its governing category by a subject, as shown in (225) (Manzini & Wexler 1987, p. 431):

- (225) [<sub>V</sub> John-wa<sub>i</sub> Mary-ni<sub>j</sub> zibun-no<sub>i/\*j</sub> syasin-o mise-ta].  
           John      Mary      Refl      pictures showed  
           “John showed Mary pictures of Refl.”

Second, while «the English pronominal *he* must be free in its governing category from subjects and nonsubjects alike», Icelandic *hann* requires to be free in its governing category by subjects, but not by objects (*ibidem*). This is shown in (226) (*ibidem*):

- (226) a. [<sub>V</sub> \*Jón<sub>i</sub> rakaði hann<sub>j</sub>].  
           John shaved him  
        b. [<sub>V</sub> Ég sendi Jóni<sub>i</sub> föt a hann<sub>j</sub>].  
           I sent John clothes for him

According to Manzini and Wexler, the linguistic facts in (225-226) show that Principles A and B of Binding Theory as formulated in Chomsky (1981a) are not restrictive enough. In fact, as for an anaphor it may not be enough to be bound in its governing category by an object, as in (225), likewise for a pronominal it may be enough to be free in its governing category from all subjects, as in (226). Hence, the picture emerging from the above analysis highlights that Principles A and B do not require to be fulfilled unconditionally but, according to Manzini & Wexler's terminology, the binding requirements of each single anaphor or pronominal must be respectively addressed in relation to some *proper antecedents* to these elements. This latter requirement takes the following form:



- (227) A. An anaphor is bound in its governing category by a proper antecedent.  
 B. A pronominal is free in its governing category from all proper antecedents.  
 (Manzini & Wexler 1987, p. 431)

As «in the concrete examples considered so far, anaphors, when bound at all, are bound by a subject; and pronominals, when free at all, are free from subjects», Manzini & Wexler propose to parametrize the notion of proper antecedent by allowing it to switch between two values, one corresponding to the subject and the other to any element (Manzini & Wexler 1987, p. 430):

- (228) A proper antecedent for  $\alpha$  is  
 a. a subject  $\beta$ ; or  
 b. any element  $\beta$ .  
 (Manzini & Wexler 1987, p. 431)

Under the revised binding principles in A and B in (227), by combining the governing category parameter in (224) with the newly proposed proper antecedent parameter in (228) «the English data, the Japanese data in [225], and the Icelandic data in [226] all correctly follow» (Manzini & Wexler 1987, p. 432). First, assuming with Manzini and Wexler that English *himself*, *each other*, and *he* are uniformly associated with value (b) of the proper antecedent parameter, then *himself* and *each other* are correctly predicted to be able to have any R-expression as their antecedent, while *he* has simply to be free from all elements. On the other hand, as far as the anaphor *zibun* and the pronominal *hann* are concerned, the correct prediction is given by associating both of them with value (a) of the proper antecedent parameter. In fact, (225) is well-formed only provided that *zibun* is bound by a subject, not by an object. Likewise, the grammaticality of (226b) with respect to (226a) shows that *hann*, while being allowed to be bound by an object, cannot be bound by a subject (cf. Manzini & Wexler 1987, p. 431).

In conclusion, the fact that Manzini & Wexler's (1987) account of long-distance anaphora resorts to a parameterized notion of governing category whereby individual parametric values are specified on each particular lexical item rather than being associated with grammatical systems as a whole testifies to the modern character of their paper within the P&P framework.

#### 2.5.4 – The verb movement parameter

Towards the end of the Eighties, a number of linguists working within Generative Grammar began to explicitly focus their attention on the study of functional categories in the inflectional domain. As briefly hinted at above in reviewing Rizzi's (1982) formulation of the Null Subject Parameter, Chomsky (1986a) is certainly fundamental in this sense in having extended the binary, X-bar-type structural schemata previously attributed only to the projections of lexical categories also to the phrasal projections of INFL and COMP, with the former being immediately dominated by the latter in the upper-left part of the sentence's phrase marker. In line with the present work's aim, it is therefore interesting to see that this enterprise also led to some important developments in the P&P framework. A seminal study for this field, Jean-Yves Pollock's article *Verb movement, universal grammar, and the structure of IP* (1989) aimed not only at shedding some light on the structural nature of the inflectional domain but, from a cross-linguistic perspective, also at providing an explanatory adequate account of some facts about the comparative analysis of French and English.

According to his own words, Pollock's working hypothesis is that «Infl(ection) should not be considered as one constituent with two different sets of features ([± Tense, ± Agr])» but that, instead, «each of these sets of features is the syntactic head of a maximal projection» (Pollock 1989, p. 365). Although this idea had been independently argued for by Moro (1988) by analyzing Italian copular sentences, Pollock's hypothesis is argued for on the basis of a collection of empirical data concerning the position of inflected verbs in English and French with respect to some elements which are assumed to always occur in the same D-Structure positions in both languages, assuming their common D-Structure form in (214) (Pollock 1989, p. 366):

(229) [<sub>IP</sub> NP I ([<sub>Neg</sub> not/pas]) [<sub>VP</sub> (Adv) V ... ]]

In the first series of tensed clauses shown in Pollock (1989), it can be easily seen that while in English the main verb cannot occur neither before the negative particle *not* nor adverbs like *often*, in French the main verb may appear higher than both the negative particle *pas* and the adverb *souvent* (Pollock 1989, p. 367):

- (230) a. \*John likes not Mary.  
           b. Jean (n') aime pas Marie.
- (231) a. \*Likes he Mary?  
           b. Aime-t-il Marie?
- (232) a. \*John kisses often Mary.  
           b. Jean embrasse souvent Marie.  
           c. John often kisses Mary.  
           d. \*Jean souvent embrasse Marie.

In previous work (Emonds (1978)) it was proposed that, contrary to English, French lacks “TENSE movement”, a specific rule lowering INFL to the verb in the VP (cf. Emonds 1978, p. 163). According to Emonds (1978), however, French grammar features an obligatory transformational rule absent in English, namely, “Finite Verb Raising”, «by which the verb of a clause with TENSE is attracted to pretense position» (Emonds 1978, p. 168). While basically accepting the latter idea, Pollock points out a number of problems facing Emonds’s analysis, including both some questions concerning the theoretical status of these rules and, regarding the aim of the present work, some important empirical facts. For example, as noted by Pollock, «French and English behave similarly when Verb Movement applies to the same elements in the two languages, namely, *be/être* and (auxiliary) *have/avoir*», as shown in (233) (Pollock 1989, p. 370):

- (233) a. He hasn’t understood./Has he understood?  
           b. Il (n’) a pas compris./A-t-il compris?  
           c. He is seldom satisfied./They are all satisfied.  
           d. Il est rarement satisfait./Ils sont tous satisfait.

Although in Emonds (1978) English and French are set apart basically in terms of a raising rule in the former language and lowering rule in the latter, according to Pollock there also exists, on a deeper and more explanatory level, an abstract difference between auxiliaries, which unanimously undergo Verb Movement in both languages, and lexical verbs, whose behaviour in this respect admits cross-linguistic variation.

The two languages also behave similarly regarding the application of Verb Movement in infinitives. As far as infinitives with auxiliary verbs are concerned, in French «*être* can but need not move to [- finite] Infl», as in (234) (Pollock 1989, p. 373):

- (234) a. Ne pas être heureux est une condition pour écrire des romans.  
 b. N'être pas heureux est une condition pour écrire des romans.

Although with a different degree of acceptability with respect to (234), the same appears to be true also for its English translation in (235) (Pollock 1989, p. 376):

- (235) a. Not to be happy is a prerequisite for writing novels.  
 b. ?To be not happy is a prerequisite for writing novels.

Considering infinitives with lexical verbs, French and English follow the very same lexical restrictions. As shown in (236-237), Pollock's data shows not only that French does not allow infinitive verbs like *sembler* to precede negation (Pollock 1989, p. 374):

- (236) a. Ne pas sembler heureux est une condition pour écrire des romans.  
 b. \*Ne sembler pas heureux est une condition pour écrire des romans.

but also that, crucially, the same is true for English infinitive verbs like *seem* (Pollock 1989, p. 376):

- (237) a. Not to seem happy is a prerequisite for writing novels.  
 b. \*To seem not happy is a prerequisite for writing novels.

In addition to supporting the existence of a deep-seated link between Verb Movement and the finiteness of the verb, which can be seen from the fact that, in all the examples seen so far from both French and English, this syntactic process appears to be obligatory only in tensed clauses, the ungrammaticality of (236b) and (237b) indicates, according to Pollock, that «Verb Movement to Infl is impossible in infinitives with lexical verbs» (Pollock 1989, p. 378). In the light of other data, however, this latter generalization seems to be falsified by the parallel grammaticality, in French, of sentences pairs like the following (Pollock 1989, pp. 377-378):

- (238) a. A peine parler l'italien après cinq ans d'étude dénote un manque de  
 don pour les langues.

b. Parler à peine l'italien après cinq ans d'étude...

“To hardly speak Italian after five years of study denotes a lack of gift for languages.”

As pointed out by Pollock, the fact that in (238b) *parler* occurs before *à peine* runs counter to the assumption that non-finite lexical verbs cannot move to Infl. Curiously, it is interesting to see that the unacceptability of (236b) – namely, the impossibility for a lexical verb to follow negation in French – is still not called into question, as remarked in the following example (Pollock 1989, p. 379):

(239) \*Ne comprendre pas l'italien apres cinq ans d'etude...

In order to keep the previous generalization, as its scope advantageously applies to both French and English instances of Verb Movement, Pollock's solution is to postulate the existence of «a Verb Movement rule, different from Verb Movement to Infl, moving the nonfinite verb to some intermediate position before the negative adverb *pas*» (*ibidem*). While in French this kind of “short” Verb Movement is not lexically restricted, in English it applies only to auxiliaries, as shown in (240) (Pollock 1989, p. 382):

- (240) a. I believe John to often be sarcastic.  
b. I believe John to often sound sarcastic.  
c. ?I believe John to be often sarcastic.  
d. \*I believe John to sound often sarcastic.

After observing that in both French and English finite contexts short Verb Movement, being lexically unrestricted in the former language but not in the latter, follows exactly the same lexical restrictions as Verb Movement to Infl, Pollock's decisive step in accounting for these apparently unrelated syntactic properties is to hypothesize that the availability of Verb Movement to Infl follows from the availability of short Verb Movement:

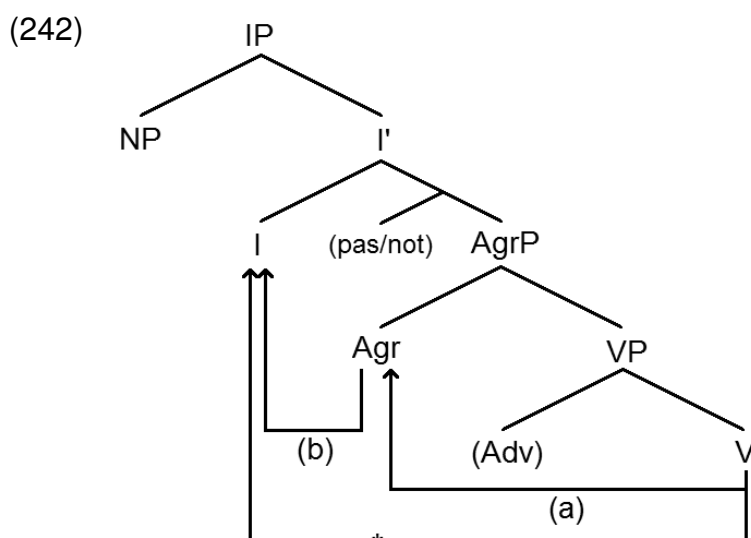
Although the French/English contrast with respect to Verb Movement to Infl ceases to exist in infinitives – in both languages only auxiliaries can undergo the rule – the very same contrast crops up in infinitives in another form: short Verb Movement is free in French but restricted to *be* and *have* in English. [...] Let us take this correlation seriously and assume

that Verb Movement to Infl exhibits lexical restrictions in tensed clauses if and only if short Verb Movement is also lexically restricted. (Pollock 1989, pp. 382-383)

More precisely, Verb Movement to Infl has not to be regarded as a process autonomous from its shorter counterpart, but rather as «the sum of two more "local" processes, the first one consisting of short Verb Movement, the second one moving the verbs to Infl from the intermediate position they thus reach» (Pollock 1989, p. 383). In these terms, starting from the assumption that Verb Movement, a canonical instance of head movement, obeys the Head Movement Constraint originally formulated in Travis (1984) as in (241):

- (241) Head Movement Constraint (HMC):  
 An  $X^0$  may only move into the  $Y^0$  which properly governs it.  
 (Travis 1984, p. 131)

and which in turn implies that the syntactic position targeted by short Verb Movement must correspond to an intermediate functional head between INFL and V, Pollock assumes that «short Verb Movement is in fact Verb Movement to Agr», that is, «a category in its own right, to be distinguished from Tense, which is the head of what has so far been called Infl», as shown in (242) (Pollock 1989, pp. 383-384):



Returning now to the lexical restrictions affecting Verb Movement, the fact that in certain configurations only auxiliaries are always allowed to move upwards (to Agr or to Tense) while lexical verbs are not can be accounted for, according to Pollock, by

appealing to the interaction between these two verb classes' respective thematic properties and a language specific, binary property of the Agr head. On the one hand, differently from lexical verbs, *have/avoir* and *be/être* are not theta-role assigners and therefore are, *ceteris paribus*, free to move to functional positions blocking theta-role assignment without triggering a violation of the Theta-Criterion, a universal constraint on X-bar theory formulated in Chomsky (1981a) and requiring that that all theta-roles be assigned:

(243) Theta-Criterion:

Each argument bears one and only one  $\theta$ -role, and each  $\theta$ -role is assigned to one and only argument.

(Chomsky 1993 [1981a], p. 36)

On the other hand, the functional head Agr differs, in English and French respectively, exactly in its capacity to either block or not theta-role assignment of the verb adjoined to it – a property which in turn is attributed by Pollock to the relative richness of the agreement inflections carried by finite verbs in the two languages:

More precisely, let us suppose that Agr in English, unlike Agr in French, is not "rich" enough morphologically to permit transmission of the verb's  $\theta$ -role(s)-in other words, that it is "opaque" to  $\theta$ -role assignment, unlike French Agr, which, being richer morphologically, is "transparent" to  $\theta$ -role assignment. (Pollock 1989, p. 385)

In these terms, the impossibility of performing Verb Movement (to Agr or to Tense) of a lexical verb in English follows from the fact that the lexical verb, by moving to an Agr functional position which is not rich enough to be transparent to theta-role assignment, would be no more able to theta-mark its arguments, thus triggering a violation of the Theta-Criterion (cf. *ibidem*).

Pollock's notion of opacity is also assumed to account for both English Affix Movement rule (the "TENSE Movement" in Emonds's (1978) terminology) and the impossibility of having Verb Movement to Tense with non-finite lexical verbs. Differently from Verb Movement, which creates a structure of the form (244) (*ibidem*):

(244) [<sub>Agr</sub> [V] Agr]

which blocks theta-role assignment, the syntactic configuration yielded by lowering opaque Agr to V via Affix Movement, which is shown in (245), allows the former to assign its theta-roles despite its opacity (Pollock 1989, p. 386):

(245) [<sub>V</sub> V Agr]

As far as Verb Movement to Tense is concerned, the result of the successive incorporation of V into Agr and T(ense) creates a structure which is similar to (244) (*ibidem*):

(246) [<sub>T</sub> [<sub>Agr</sub> V Agr] T]

but that, crucially, never allows the adjoined V to head a theta-chain in infinitives since, according to Pollock's hypothesis, «[-finite] Tense is "opaque" to  $\theta$ -role assignment (perhaps universally)» (*ibidem*).

Summing up so far, Pollock's «"opacity" versus "transparency" parameter» provides an account for the cross-linguistically different lexical restrictions concerning Verb Movement in terms of theta-theory violations (Pollock 1989, p. 391). However, another theoretical aspect which needs to be explained consists in the reason why, as shown in (232d) (here repeated as (247)), Verb Movement is obligatory – that is, it cannot be replaced by Affix Movement – in French tensed clauses (cf. Pollock 1989, p. 372):

(247) \*Jean souvent embrasse Marie.

As it will soon be seen, a solution for addressing this problem is hinted by the fact that, as previously seen in (227-230), «in English and French infinitives, *have/avoir* and *be/être*, although they can move to (Agr and to) Tense, need not do so», that is, the same generalization suggesting the existence of a link between Verb Movement and [+ finite] Tense (Pollock 1989, p. 392). This link, according to Pollock, corresponds to the quantificational nature of inflected Tense, as formulated in (248) (*ibidem*):

(248) [+ finite] Tense (that is, [ $\pm$  Past]) is an operator.

Starting from the assumption that «[ $\pm$  Past] will have to bind a variable, like other operators», Pollock's hypothesis is that, on the one hand, the application of Verb



Movement to Infl in tensed clauses is parallel to wh-movement in leaving a trace corresponding to a variable but, on the other hand, only the empty V position left by the former process can adequately be bound by [ $\pm$  Past], as expressed in the following definition (Pollock 1989, p. 392):

(249)  $\alpha$  is a variable for [ $\pm$  Past] iff  $\alpha = [_{Vi} e]$  bound by [ $\pm$  Past].

Seen in this light, (247) is due to the fact that if Verb Movement does not apply in the tensed clause, then there is no variable which can be bound by the [ $-$  Past] feature on Tense and hence «the ban on vacuous quantification in natural languages [...] excludes it» (Pollock 1989, p. 393). Moreover, the fact that also sentences derived by short Verb Movement, like (250), are ungrammatical (*ibidem*):

(250) \*Pierre ne pas mange souvent.

[<sub>TP</sub> NP ne [<sub>T</sub> – Past] pas [<sub>Agr<sub>i</sub></sub> [<sub>V<sub>i</sub></sub> mang- Agr]] [<sub>VP</sub> Adv e<sub>i</sub>]]

shows in turn that [ $\pm$  Past] can only bind the verb's trace by directly inheriting its index, namely, by means of the incorporation of V into Agr and T through successive cyclic X<sup>0</sup> movement, as shown by the well-formedness of (251) (*ibidem*):

(251) Pierre ne mange pas.

[<sub>TP</sub> NP ne [<sub>T<sub>i</sub></sub> [<sub>Agr<sub>i</sub></sub> [<sub>V<sub>i</sub></sub> mang] - Agr]] – Past] pas e<sub>i</sub> [<sub>VP</sub> e<sub>i</sub>]]

As predicted by Pollock's theory, no problems arise in this respect with French lexical infinites, as [ $-$  finite] Tense is expected to be universally opaque to theta-role assignment and to not involve quantification. Therefore, the grammaticality of (252) and the ungrammaticality of (253) are fully accounted for (Pollock 1989, p. 394):

(252) Ne pas manger...

[<sub>TP</sub> PRO ne e<sub>i</sub> pas e<sub>j</sub> [<sub>VP</sub> [<sub>V</sub> mang- [<sub>T<sub>i</sub></sub> – finite] + Agr<sub>j</sub>]]]

(253) \*Ne manger pas...

Although the analysis so far effectively accounts for the obligatoriness of Verb Movement to [ $+$  finite] Tense in French, the same pattern of results cannot be reconciled

with available evidence from English. This problem can be easily seen in the following pair of English sentences (see also (232c) above) (Pollock 1989, p. 395):

- (254) a. John left.  
b. \*John not left.

If, in order to justify Affix Movement in (254a), we postulated that inflected Tense does not correspond to a variable-binding operator in English, on the other hand «we would also expect [254b] to be well-formed (recall that French *Ne pas manger...*, *Ne pas partir ...* are fine)» (*ibidem*). Therefore, the parameterization of the quantificational nature of [ $\pm$  Past] would produce contradictory results. Furthermore, the fact that in (254b) the feature [ $-$  Past] requires the lexical verb to move to (Agr and to) Tense but, conversely, English Agr is a thematically opaque position, seems to indicate that for languages like English there must be some other option which is not available for French. As noted by Pollock, one possible option in this sense consists in «allowing an auxiliary verb generated beyond the VP barrier to count as a substitute for the immovable main verb in the VP» (Pollock 1989, p. 396). More precisely, his hypothesis is that this auxiliary verb, which is assumed to be generated under Agr and subsequently moved to [ $+$  finite] Tense, corresponds to English *do*, as shown by examples (255) and their derivation (256) (Pollock 1989, p. 399):

- (255) a. John did not go.  
b. John doesn't understand.  
(256) [<sub>TP</sub> John [<sub>Ti</sub> [<sub>Agr<sub>i</sub></sub> [<sub>Vi</sub> do ] Agr] T ] [<sub>NegP</sub> not] [<sub>AgrP</sub> e<sub>i</sub> [<sub>VP</sub> V]]]

In (256), *do* satisfies the requirement that its trace be bound by [ $\pm$  Past] as it represents a copy of the theta-grid of the lexical verb. This position, which Pollock himself addresses as actually rather traditional, is argued on the assumption that *do* does in fact substitute the whole VP in sentences like «*John took more time over it than he had ever done before*» or «*I don't know the answer but Peter may do*» (*ibidem*). Crucially, however, there are some kinds of VPs which cannot be substituted by *do*. For example, this is true of any VPs containing a copula and an adjective, as in (257) (Pollock 1989, p. 400):

- (257) a. \*John doesn't be happy.  
 b. \*John does not have gone.  
 c. \*John did not be singing.  
 d. \*John didn't be kissed by Mary.

Returning now to more simple affirmative sentences such as (254) (here repeated as (258a)) and (258b) (Pollock 1989, p. 404):

- (258) a. John left.  
 b. John leaves.

their Affix Movement analysis can be argued by assuming they correspond to (259a-b) (*ibidem*):

- (259) a. John did leave.  
 b. John does leave.

Pollock's idea is that, along with lexical *do*, English had also developed a non-overt auxiliary  $\emptyset$  which features the very same thematic properties:

[...] I will assume that English has a nonlexical counterpart of *do*, call it  $\emptyset$ , which shares with it all its defining properties except its lexical character. In particular,  $\emptyset$ , like *do*, is a substitute verb and can therefore copy the  $\theta$ -role of the main verb in the VP. Like *do*,  $\emptyset$  is generated under Agr and moves to Tense. (*ibidem*)

Being the actual non-lexical counterpart of *do*,  $\emptyset$  is generated in the same base position and moves to [+ finite] Tense in order to satisfy the latter's quantificational properties. The derivation of (258) is thus shown in (260) (*ibidem*):

- (260) [<sub>TP</sub> John [<sub>Ti</sub> [<sub>Agr</sub> [<sub>Vi</sub>  $\emptyset$  ] Agr] T ] [<sub>AgrP</sub> e<sub>i</sub> [<sub>VP</sub> leave]]]

Now, since  $\emptyset$  alternates freely with *do*, Pollock observes that nothing would rule out the derivation of the following ill-formed sentences and their S-structure (262) (Pollock 1989, p. 405):

(261) a. \*John not left.

b. \*John not leaves.

(262) [TP John [T<sub>i</sub> [Agr<sub>i</sub> [V<sub>i</sub>  $\emptyset$ ] Agr] T ] [NegP not] [AgrP e<sub>i</sub> [VP leav-]]]

In fact, the ungrammaticality (261) has to do with the independent fact that, according to Pollock, «NegP is a maximal projection and, unlike AgrP, an inherent barrier» which «must be L-marked to become transparent to antecedent-government» between Tense and the original position of the non-overt auxiliary  $\emptyset$  (Pollock 1989, p. 405). Within the *Barriers* framework, “L-marking” corresponds to the specific syntactic relation which holds between a lexical head and its sister node in terms of X-bar theory (Chomsky 1986a, pp. 13-15). Consequently, since in (262)  $\emptyset$  is not lexical, NegP is not L-marked when  $\emptyset$  (together with Agr) moves to its sister node T. Therefore, this maximal projection acts as a barrier and triggers a violation of the ECP (cf. Pollock 1989, p. 405).

Finally, the very last part of Pollock’s paper deals with the impossibility of using the non-overt auxiliary  $\emptyset$  in declaratives like *John is happy*, whose derivation is showed below (Pollock 1989, p. 406):

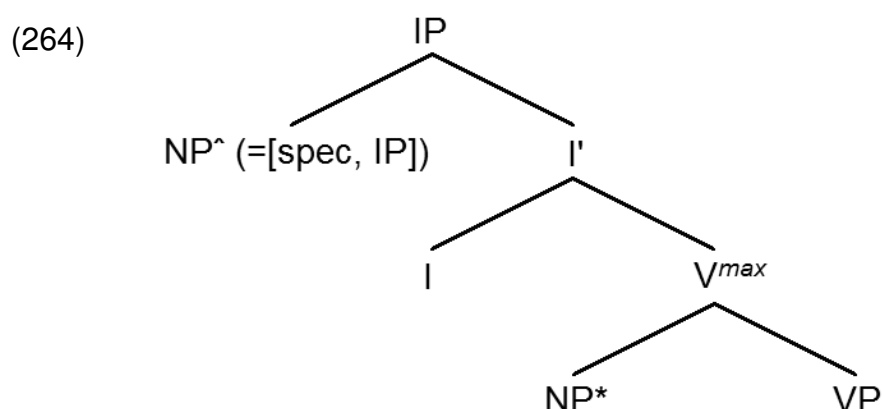
(263) [TP John [T<sub>i</sub> [Agr<sub>i</sub> [V<sub>i</sub>  $\emptyset$ ] Agr] – Past] [AgrP e<sub>i</sub> [VP be A]]]

While abiding by the ECP and the conventions requested for the co-indexation of T and the trace left by  $\emptyset$ , the obligatoriness of Be/Have Movement in declaratives follows from the fact that the auxiliary *do*, whether lexical or non-lexical, cannot thematically substitute any VPs containing a copula and an adjective as already pointed out for (256).

In conclusion, Pollock’s decomposition of the inflectional domain into T(ense)P and Agr(eement)P and his subsequent parameterization of the Agr head as either opaque or transparent to theta-role assignment is a perfect example of attributing a set of apparently independent cross-linguistic differences to a syntactic property varying in a binary fashion. In addition to this, the fact that this “opacity versus transparency” parameter is based on the feature specification of a functional head makes this work, together with Rizzi (1982), one of the most up-to-date parametric inquiries of the GB Theory framework.

### 2.5.5 – The parameter of nominative case assignment

Along those parametric studies which, like Pollock (1989), provided insights on the basic structure of the sentence which were beyond reach with approaches earlier than Chomsky's (1986a) *Barriers* framework, another important paper in this sense is *The position of subjects* (1991) by Hilda Koopman and Dominique Sportiche. In this work, Koopman & Sportiche's proposal was that, assuming the existence of a universal structural position in which the subject of the main verb is generated, languages can be divided into two basic classes according to whether the subject does, or does not, move from such base position to the specifier of the next higher functional head, that is, INFL, as shown in the following phrase marker (Koopman & Sportiche 1991, p. 212):



In (264), NP\* represents «the canonical or D-structure position of the subject», while NP^ corresponds to «its S-structure position in simple declarative clauses» in languages like English (*ibidem*).

According to Koopman & Sportiche, the fact that in English the subject moves from its base position to the specifier of INFL has to do with the raising nature of tensed INFL. By comparing the prototypical raising verb *seem* with modals, which are in turn assumed to appear in INFL, it is argued that there are some striking similarities between the two, syntactically as well as semantically. From a syntactic point of view, the verb *seem* does not assign an external theta role. This property is in turn regarded as one, if not the main, syntactic factor in determining the raising nature of a category, as from the lack of a syntactically projected external argument a further set of properties generally attributed to raising categories immediately follow. Such properties are listed in (265) (Koopman & Sportiche 1991, p. 213):

- (265) a. *seem* imposes no selectional restrictions on its subject;  
 b. *seem* can take expletive *it* as a subject or non-expletive subjects;  
 c. *seem* allows as subject an NP licensed by the predicate of the clause embedded under it;

Although (265) are equally important, (265b-265c) are the properties which are attributed more attention by the authors. In fact, the possibility for raising verbs to take as subject either an expletive *it*, as in (266a), or an NP licensed by the predicate of the lower clause, as in (266b), supports the fact that «there is another, less straightforward reason», this time semantic in nature, «why *seem* is treated like a raising verb» (Koopman & Sportiche 1991, p. 213):

- (266) a. It seems that John sleeps all day.  
 b. John seems to sleep all day.

As noted by Koopman & Sportiche, the fact that «the relevant syntactic representation of [266], when looked at appropriately, is in fact *seems John to sleep all day* [...] i.e. essentially identical to the first sentence» suggests another crucial semantic property shared by raising categories (Koopman & Sportiche 1991, p. 214). As *seem* takes as its complement a full proposition, which in turn corresponds to a saturated predicate, according to the authors all raising categories including those actually not allowing an expletive subject and a sentential complement will, by extension, take a saturated predicate as complement. This property is formalized in (267) (Koopman & Sportiche 1991, p. 215):

- (267) No category takes as complement a syntactic category corresponding to a non-saturated predicate.

Turning now to modals, according to Koopman & Sportiche «it appears natural to suppose that they take saturated predicates or propositions as arguments» (*ibidem*). In fact, as shown in the following example (*ibidem*):

- (268) a. Mary might sleep all day.  
 b. It is possible that Mary sleeps all day.

the modal *might* in (268a) has its semantic import expressed by the main clause in (268b), which in turn takes as its complement the saturated predicate corresponding to the non-INFL material of the former sentence. Moreover, by assuming that «there is as transparent a correspondence as possible between “semantic constituent structure” and syntactic structure» in such a way that, «if INFL takes a saturated predicate as argument, it is a priori reasonable to postulate that this semantic fact is reflected syntactically», not only does this position further equate tensed INFL to raising verbs semantically but, from a syntactic point of view, it is also reasonable to assume that the NP subject and the VP predicate constituting the argument of INFL form a single syntactic constituent, which is represented in (264) by the projection  $V^{\max}$  (Koopman & Sportiche 1991, p. 215).

In order to argue for the raising nature of tensed INFL, a comparison is made by Koopman and Sportiche between the arguments previously put forth in (265) for analyzing the verb *seem* as a raising verb and some syntactic properties of the modal *will*, which is assumed to represent all categories generated in tensed INFL. These properties are listed in (269) (Koopman & Sportiche 1991, p. 216):

- (269) (i) *will* does not assign an external theta role;  
 (ii) *will* allows as subject an NP licensed by the predicate embedded under it;

By indicating (269) as «diagnostic properties of raising items», Koopman & Sportiche determine the raising nature not only of *will* but, by extension, of all the modals, here regarded as the set of all the categories appearing in tensed INFL (*ibidem*).

By the same arguments all the modals are raising categories, *do* is a raising verb, and more generally tensed INFL is a raising category [...]. By the same argument, aspectual verbs (perfective *have* and *avoir*, passive *be* and *être*, progressive *be*), which are analyzed as heading their own VP and taking VP complements are raising verbs. (*ibidem*)

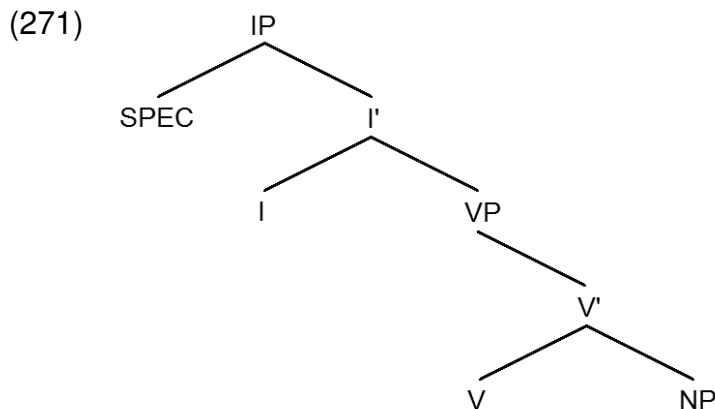
Another determining factor supporting both the definition of tensed INFL as a raising category and the analysis proposed in (264) concerns the syntactic relation between NP\* and VP. In this respect, Koopman & Sportiche's central observation is that, as «lexical relations such as selection, subcategorization or theta assignment can hold only of items that are structurally close», theta assignment requires the theta-marker and the target of theta-marking to be sister nodes (Koopman & Sportiche 1991, p. 213). Far from being an unprecedented claim, this proposal dating back to Chomsky (1986a) does in fact require a

slight but crucial adjustment precisely in the case the theta-marker and the target of theta-marking correspond to the VP and the NP subject respectively. As can easily be seen in (270) (Koopman & Sportiche 1991, p. 217):

(270) [<sub>IP</sub> [<sub>spec</sub>, I] [<sub>I</sub> I [<sub>VP</sub> [<sub>V</sub> V NP]]]]

«sisterhood must be so defined as to ignore intermediate projections of INFL» in order to allow NP\* to select VP as its complement (*ibidem*). However, by assuming that INFL is a raising category selecting  $V^{max}$ , sisterhood between NP\* and VP would directly follow, thus allowing the general theory of theta relations to avoid the undesirable complication represented by such an *ad hoc* amendment as the one postulated for structures like (270).

Generalizing to other languages the assumption that NP\* and VP are sister nodes while INFL and NP\* are not, in Koopman & Sportiche (1991) some more arguments are provided also with respect to syntactic order and subject-verb agreement. First, considering a uniform VSO language like Welsh or Irish, by assuming the correctness of (270) (here repeated as 271):



it follows that the VSO order cannot be base generated as such, the reason being that the subject is generated as the specifier of INFL, that is, outside the constituent VP which includes V and the object O. Assuming, therefore, that a VSO structure involves syntactic movement of V to a structural position higher than SPEC-INFL, the observation that both Welsh and Irish are AuxSVO when an auxiliary appears suggests that «the VSO structure involves movement of the V to the position that an overt Aux otherwise occupies», which in turn could *a priori* correspond to either COMPL or INFL (Koopman & Sportiche 1991, p. 219). Regarding the first possibility, as according to Den Besten (1983) V-to-C movement



cannot take place if the corresponding complementizer is overt, the fact that in Welsh and Irish the VSO order is allowed also in clauses containing a lexically filled complementizer supports the idea that such order involves movement of V to INFL. At this point it follows not only that the previously alluded AuxSVO structures are actually base generated, but also that, if then it is assumed that the structure of Irish and Welsh is INFL SVO, structure (270-271) cannot directly account for this base generated order while (264) can (cf. Koopman & Sportiche 1991, p. 220).

After Welsh and Irish, Koopman and Sportiche's analysis turns to Standard Arabic. According to the authors, what is interesting about this language is its agreement pattern variation in relation to word order, which can optionally be SVO or VSO. More precisely, while in SVO structures this language exhibits full agreement with the subject, on the other hand «in the order VSO [...] the verb only exhibits a default number agreement» (Koopman & Sportiche 1991, p. 221). As agreement mirrors a syntactic relation between a head and its specifier, in this specific case agreement is morphologically realized on the verb as the latter is moved into INFL. In these terms, with the SVO order the verb fully agrees with the subject because, after the verb moves into INFL, the subject moves to SPEC-INFL and therefore the verb has a lexical specifier with which it can agree in all its features. In the case of the VSO order, on the other hand, since there is no subject movement to the specifier of INFL after the verb has moved into INFL «INFL has a silent expletive specifier [...] and agreement gets the default value, namely 3rd person singular» (*ibidem*).

By comparing the empirical data observed in Irish/Welsh and Standard Arabic with their initial analysis of tensed INFL in English as a raising category, Koopman and Sportiche's conclusion is that what forces the subject to raise from its D-structure position to the specifier of INFL – that is, from NP\* to NP<sup>̂</sup> in (264) – in some languages but not in others is a difference in nominative Case assignment. In fact, similarly to what happens in English with passives and raising-to-subject constructions, if a lexical NP subject needs Case but its base position is not Case-marked, it can acquire Case provided it moves to a higher Case-marked position:

This is reminiscent of the of the obligatory character of NP-movement in passive constructions in English or in raising constructions, suggesting a characterization of this effect in terms of Case theory. (Koopman & Sportiche 1991, p. 227)

In this terms, «when NP\* is a Caseless position, an NP in it which needs Case must move» but, conversely, «if on the other hand NP\* is a Case position, movement is not necessary» (Koopman & Sportiche 1991, p. 227). Building on this intuition, the authors propose that Case can be assigned not only under government by a Case assigner, as traditionally assumed, but also under agreement with a Case assigning head, with agreement here meant as the syntactic relation between a head and its specifier. (cf. Koopman & Sportiche 1991, p. 229). In the case at hand, starting from the assumption that [+Tense] INFL assigns nominative Case, according to Koopman & Sportiche «INFL can assign Case by agreement or governed Case, depending on the language» (*ibidem*). This distinction explicitly goes back to the familiar opposition between structural case and inherent case, with the former expressing the configurational property whereby an NP is assigned nominative/accusative Case according to the structural position it occupies and the latter being also thematically dependent, but with an important difference: if on the one hand in Chomsky (1981a) structural case was assumed to be assigned only under government, in Koopman & Sportiche (1991) «nominative is both structural Case», that is, assigned by government, «and Case by agreement, while inherent Case is always governed Case» (*ibidem*).

In order to account for the government relation between INFL and NP\*, in Koopman and Sportiche (1991) the notion of head-government is extended to include not only the head's sister node – that is, in the case of INFL,  $V^{\max}$  – but also the latter's specifier:

There is a good deal of evidence (see Chomsky 1986 [=1986a, A.R.], Sportiche 1988b [=1988, A.R.], 1990) suggesting that if some  $X^0$  governs YP, it governs the specifier of YP. Since clearly I governs  $V^{\max}$  (sisterhood surely entails government), it should, by this definition govern its specifier, namely NP\*. (Koopman & Sportiche 1991, p. 228)

Then, in order to exclude government between INFL and NP<sup>^</sup>, which would otherwise overlap with the head-specifier relation represented by agreement, the following definition of government is adopted:

(272) *Government*

A governs B if A X-commands [= i-commands] B and no barrier for B intervenes between A and B. (Koopman & Sportiche 1991, p. 229)

(273) *I-Command*

A i-commands (immediate command) B if the first constituent (distinct from A) containing A contains B. (Koopman & Sportiche 1991, p. 230)

With this formal apparatus in place, Koopman and Sportiche account for the different positions of subjects in languages like English, Arabic and Irish by means of the following syntactic parameter (Koopman & Sportiche 1991, p. 232):

If it is a Case assigner, INFL is a structural Case assigner. The Case of INFL [...] varies as follows:

- (274) i. INFL is specified as a governed Case assigner or not.  
 ii. INFL is specified as an agreement Case assigner or not.

Koopman and Sportiche's parameter in (274) yields the four possible types of languages which are shown in the following table (*ibidem*):

(275)

	+ Governed Case	– Governed Case
+ Agr Case	Arabic Finite Clause	French Finite Clause English Finite Clause Portuguese Inflected Infinitive
– Agr Case	Irish Finite Clause Welsh Finite Clause	English Infinitive French Infinitive

In those languages, like French and English, in which INFL is specified as an agreement Case assigner but not as a governed Case assigner, NP\* receives no Case. As a consequence, in order to avoid the Case Filter the external argument cannot surface as NP\* but has to raise to SPEC-INFL, from which position subject agreement is triggered. In languages such as Arabic, in which INFL assigns Case under both government and agreement, the subject can be realized in either NP\* or NP^ position, with morphological agreement triggering only in the latter case. If INFL is specified as a governed Case assigner only, there is no reason for the subject to raise from NP\* to the specifier position of INFL, as the position in which it is base-generated is already a Case marked position. According to the authors, this description applies to Irish, a uniformly VSO language in which «the Subject never agrees with the verb or the auxiliary in INFL» (Koopman & Sportiche 1991, p. 231). The remaining possibility in Koopman and Sportiche's table above applies to the external arguments in non-finite clauses. Since [–Tense] INFL is not a

Case assigner and, by extension, there is no nominative Case available in infinitives, «no overt subject or covert subject requiring Case can surface there» and therefore the external argument has to rise to a higher Case-marked position, as in raising-to-subject constructions (Koopman & Sportiche 1991, p. 235).

In conclusion, the parametrization of the INFL head with respect to Case assignment proposed in Koopman & Sportiche (1991) represents another important example of generative grammarians' commitment to reduce a set of typological differences, in this particular case essentially related to the different surface position of subjects and the possibility of subject agreement, to the possibility of having cross-linguistically different specification of a given functional head.

## 2.6 – Conclusion

After reviewing the main syntactic parameters of the GB Theory, some relevant aspects seem to emerge with respect to both the classification proposed in section 1.1 and the overall development of the notion of parameter. First of all, although our preliminary classification appears to generally hold true for the typological phenomena addressed by the classes of Locality, Merge, and Linearization parameters, the results of this comprehensive review also show us that the Case assignment parametric class needs to be partially modified by shifting the so-called *P-stranding* parameter into the class of Spellout parameters. In fact, since according to Baker (1988) V-P reanalysis is to be ascribed to abstract incorporation, its associated phenomena should in turn be analyzed in terms of incorporation and, therefore, as an instance of head-movement, here meant as the epiphenomenal consequence of the application of Internal Merge to an  $X^0$  element and the subsequent deletion of the latter's remerged copy. Second, despite Chomsky's (1981b) acknowledgement that parameters are not directly expressed on universal principles but rather interact with their syntactic effects at a more superficial level, it is particularly interesting to observe that in the second half of the Eighties there still was no unique nor precise format for the formulation of parameters. In fact, if on the one hand Rizzi's (1982) formulation of the null subject parameter as a binary feature [ $\pm$  pronoun] specified on INFL was strikingly ahead of its time, on the other hand the syntactic parameters proposed in Hale (1983) and in Baker (1988) still referred to variation in universal principles; and this without considering the proposal put forth in Huang (1982), according to which a set of systematic cross-linguistic differences could be attributed to the

possible variation in the locus of wh-movement. Of course, it must not be overlooked that the majority of the syntactic parameters which were formulated from the mid-Eighties onwards more or less adhered to the same lexical-based format as that of the null subject parameter; and which, as will be discussed in more detail in Chapter III, was precisely the format upheld by Borer's restrictive hypothesis on language variation. However, if on the one hand whether such a lexically based conception of language variation did hold for all parameters was still an open issue within the GB framework, with the advent of the Minimalist Program this hypothesis would become a central matter of debate in Generative Grammar.



## Chapter III

### The development of the concept of parameter in Minimalism

#### 3.1 – Introduction

Since its systematization in the early Eighties, the concept of parameter has been a constant presence through the development of the Chomskian program. Despite the profound evolution of Generative Grammar's conceptual framework, the idea that language faculty provides an innate architecture composed by universal properties common to all languages and, at the same time, only permits restricted patterns of variation, still largely underlies current generative research. In fact, the Principles and Parameters (P&P) approach seems to be not only a viable solution to the so-called logical problem of language acquisition, but also «the consensus view of the overall structure of the language faculty» (Hornstein *et al.* 2005, p. 5).

Despite the above premises, however, to say that after more than thirty years the parametric approach still counts as the standard model in Generative Grammar does not mean that there is absolute unanimity on the validity, both theoretical and empirical, of this working hypothesis. This fracture has become particularly evident in the period following the systematization of the Minimalist Program (MP), whose primary aim has been to investigate to what extent the human language faculty represents «an optimal solution to minimal design specifications, conditions that must be satisfied for language to be usable at all» (Chomsky 2001b, p. 1). Given the fact that one core tenet of this program is that «language is an optimal solution to such conditions», as stated by the Strong Minimalist Thesis (SMT), today's generative research has raised a fundamental tension between the minimalist pressure towards language faculty's structural simplicity and the idea of an overspecified P&P-like architecture – a tension which resulted, in some cases, in the calling into question of the validity of the notion of parameter itself (*ibidem*). Further complicating matters, even within parametric theory there is no firm consensus on neither the role of parameters in today's linguistic research nor the exact nature of this concept. This disagreement is in part a matter of purely methodological considerations: for instance, it is totally understandable that a particular conception of parameter can be more suited to the investigation of a specific type of linguistic phenomena than another, or that it can effectively contribute to a more elegant account of certain linguistic patterns. Nevertheless,

especially as far as the validity itself of the concept of parameter is concerned, the coexistence of such different and irreconcilable positions is due not only to different epistemological views regarding the interpretation of cross-linguistic evidence, but also to discrepancies between some basic assumptions about the architecture and functioning of language faculty itself.

Although it is certainly true that «over the last thirty years [...] the very notion of parameter has undergone significant changes, under the pressure of theoretical developments and empirical evidence», what has been happening from the 2000s onwards can be regarded as an actual crisis of parametric theory (Rizzi 2013, p. 313). Of course, this does not mean that the parametric approach to language variation is bound to be definitively rejected by Generative Grammar, especially given the success of the P&P model in resolving the tension between descriptive and explanatory adequacy (cf. Chomsky 2005, pp. 7-8). With this in mind, the current debate might as well contribute to a solution for this crisis by providing new perspectives on what is really important about parametric theory and, conversely, what should be amended of it. An analysis of the debate about the concept of parameter in the 2000s can therefore be useful to both bring its main trends of thought to light and, by doing so, morph this stalemate situation into a constructive revaluation of this notion.

### 3.2 – The parametric approach to language variation in the Government-Binding Theory

#### *3.2.1 – From parametrized principles to lexical parameters*

In the first phase of the P&P model, the original conception of parameter was based on the assumption that cross-linguistic differences directly depended on structural options specified on universal principles by Universal Grammar (UG); in other words, «it was implicitly admitted that virtually any aspect of UG could be parametrized» (Rizzi 2013, p. 313). This particular hypothesis was directly exemplified in the parametrization of Subjacency (Rizzi 1982): while on the one hand this condition consists in a universally operative principle, on the other hand its formulation is flexible enough to allow its bounding nodes to be specified each time for each specific language, thus determining different locality effects in different grammars. Very early on, however, the “classical” idea that parameters were expressed on principles began to lose ground in conjunction with the formulation of the so-called “Borer-Chomsky” conjecture (as reformulated in Baker



(2008a)), which caused the locus of parametric variation to be shifted from UG's principles to «the properties of the inflectional system», and thus to the lexicon (Borer 1984 [1983], p. 29):

All parameters of variation are attributable to differences in the features of particular items (e.g., the functional heads) in the lexicon. (Baker 2008a, p. 3)

Although having the clear advantage of representing a step forward in the development of an acquisition model for language, as «the burden of learning is placed exactly on that component of grammar for which there is strong evidence of learning: the vocabulary and its idiosyncratic properties» (Borer 1984 [1983], p. 29), in the Nineties the Borer-Chomsky conjecture also «turned out to be particularly congenial to the “light” concept of UG that Minimalism assumed» (Rizzi 2013, p. 314). In fact, the abandonment of the parametrized principles hypothesis allowed Generative Grammar to fully embrace not only the minimalist idea that «there is only one computational system» for human language, but also that this uniformity is due to its fully optimal architecture, in agreement with the Strong Minimalist Theory (Chomsky 1995a, p. 26).

### *3.2.2 – Parameters and linguistic typology in the Government-Binding Theory*

The original P&P model as presented in Chomsky (1981a) was explicitly based also on the idea that each individual parameter setting could ideally have repercussions not only for one grammatical property but also for a complex – or, according to a more modern terminology, a *cluster* – of otherwise independent linguistic properties:

Each of the systems of [the grammar] is based on principles with certain possibilities of parametric variation. Through the interaction of these systems, many properties of particular languages can be accounted for. We will see that there are certain complexes of properties typical of particular types of languages; such collections of properties should be explained in terms of the choice of parameters in one or another subsystem. In a tightly integrated theory with fairly rich internal structure, change in a single parameter may have complex effects, with proliferating consequences in various parts of the grammar. Ideally, we hope to find that complexes of properties differentiating otherwise similar languages are reducible to a single parameter, fixed in one or another way. (Chomsky 1993 [1981a], p. 6)

The idea that «when a small change is introduced there are often consequences throughout this range of phenomena», which for the first time could potentially achieve explanatory adequacy on a large scale with minimal descriptive effort, had such an immense appeal that, at the dawning of the P&P model, together with explanatory adequacy it was regarded as the ideal prerequisite of a desirable linguistic theory (Chomsky 1993 [1981a], p. 3). By means of such a theory, Generative Grammar would in fact be able not only to reconcile language universality and typological variation, but also to account for previously unexplainable clusters of grammatical properties. Moreover, any discovery in this specific direction would in turn be evidence that grammatical knowledge is acquired on a more abstract basis than what can be inferred or taught from the outside, thus further supporting the explanatory adequacy of this new theoretical approach.

The parameters that represent this very perspective on language variation have been traditionally known as «macroparameters» (Baker 1996, p. 7). The first and most influential parameter of this kind was the Pro-drop Parameter (or Null Subject Parameter), based on Rizzi (1982) and then formally formulated in Chomsky (1981a). As noted in Baker (2008a), however, «history has not been kind to the Pro-drop Parameter as originally stated» as «the cluster of properties that this parameter claimed to be related fragments in various ways when one looks at a wider range of languages» (Baker 2008a, p. 1). Moreover, the same trend towards fragmentation affected also other cases of alleged macroparameters which regarded word order, movement and configurationality variations. As a result, this way to conceptualize large-scale linguistic variation has progressively tended towards an orientation in the opposite direction, thus causing the decomposition of the original macroparameters of the GB era in terms of smaller parametric arrays and, accordingly, giving up the claim that a single dimension of variation could be associated with a cluster of typological proprieties interrelated in nonarbitrary ways (cf. Baker 1996, p. 7):

It is obvious to anyone familiar with the field that this is not what has happened. On the contrary, parameters have tended to become smaller and more construction-specific, rather than larger and more general. Medium-sized parameters have split up into “microparameters”, rather than merging into macroparameters. (*ibidem*)

Although the gradual reduction in the descriptive and explanatory scope of the concept of parameter could suggest that the parametric approach has completely lost all its original lustre by now, in today’s Generative Grammar this notion is still regarded by

many as both a resource for the conceptualization of cross-linguistic differences and the basis of language acquisition. In fact, while on the one hand the classical parameter theory of the Eighties – namely, the idea that the principles of UG are parametrized – is still being supported by some scholars, on the other hand some more concrete and practical alternatives have been proposed in order to reconcile the formal description of individual languages and the more general quest for linguistic universals with the different methodological approaches to the study of language diversity carried out in Minimalism.

### 3.3 – The parametric approach to language variation in Minimalism

As the early MP embraced the programmatic hypothesis that «in its basic structure, the language faculty has properties of simplicity and elegance that are not characteristic of complex organic systems» (Chomsky 1995a, p. 29), the traditional view that parametric options were prespecified within a highly structured theory of UG was replaced by the assumption that «typological variation should reduce to the ordering parameters and properties of functional elements» (Chomsky 1995a, p. 61). These theoretical premises, which are still considered to be the only plausible from an evolutionary perspective, have caused the inquiry into the nature of parametric variation among languages to go in two main directions. On the one hand, the mainstream minimalist view points towards a minimally specified UG with variation strictly limited to microparameters. On the other hand, other generative linguists defend the idea of an overspecified theory of language faculty by directly supporting the plausibility of concept of macroparameter.

#### 3.3.1 – *Kayne's microparametric approach to language variation*

After the realization that «it is not likely that a single parameter may neatly control a certain cluster of properties» (Rizzi 2013, p. 316), the general trend in modern parametric theory has been to reduce the scope of parameters from complexes of typological properties to finer-grained cross-linguistic differences. However, the main obstacle to such an inquiry has been represented by the difficulty of effectively relating each single parametric choice with its corresponding effects on language variation. For example, if two languages differ in the setting of more than one parameter and, in addition to this, each of these parametric differences is responsible for a complex cluster of morphosyntactic

properties, in principle it is not clear at all how to determine which linguistic differences depend on which parameter respectively:

In essence, in searching for clusters of properties, one must make decisions about what syntactic differences can plausibly be linked to what other syntactic differences. To a certain extent one is guided by one's knowledge of syntax in general and by the theory within the framework of which one is working. Such general considerations do place limits on the set of hypotheses one takes seriously, but typically the set of plausible linkings remains larger than one would like. (Kayne 2000, p. 4)

In order to better capture the deductive connections deriving from each individual parameter, the most important step in this direction consisted in taking into comparative account sets of closely related languages, thus limiting any unwanted overlaps between the visible effects of different parameter settings. This peculiar approach to comparative syntax is represented by Kayne's (2000) microparametric syntax: by comparing «historically close systems [...] which did not have the time to differentiate along too many parametric dimensions» (Rizzi 2013, pp. 316-317), this perspective allows the linguist to better isolate the effects of each parameter, thus «providing results of an unusually fine-grained and particularly solid character» (Kayne 2000, p. 5). In these terms, Kayne compares the degree of accuracy of this approach to comparative syntax to that of a controlled experiment in physical sciences, where all the variables are kept apart from the factor under study so that the effect or influence of that factor can be precisely identified and investigated in detail:

If it were possible to experiment on languages, a syntactician would construct an experiment of the following type: take a language, alter a single one of its observable syntactic properties, examine the result to see what, if any, other property has changed as a consequence. If some property has changed, conclude that it and the property that was altered are linked to one another by some abstract parameter. Although such experiments cannot be performed, I think that by examining pairs (and larger sets) of ever more closely related languages, one can begin to approximate the results of such an experiment. (*ibidem*)

Compared with the more traditional macroparametric approach, «by examining sets of very closely related languages, languages that differ from one another in only a relatively small number of syntactic ways», Kayne's microparametric syntax allows the linguist to really grasp the minimal units of linguistic variations as a natural scientist would

do by means of «a new kind of microscope with which to look into the workings of syntax» (Kayne 2005, pp. 8-9).

Rather than representing a radical shift from the P&P model of the Eighties, the notion of microparameter represents to some extent the natural development of the classical notion of parameter: while the latter's aim was to account for large clusters of different properties, and therefore ideally aimed at comparing specimens of typologically distinct grammars, the former is based on the same basic intuitions of the classical parametric approach, with the only methodological difference of being applied on a small-scale linguistic variation. As such, with the term "microparameter" Kayne identifies «those parameters that at least in some cases differentiate two very closely related languages» (Kayne 2005, p. 7). In this sense, an example of microparameter is represented by the difference between the Italian complementizer *se* and its French counterpart *si*. In Italian, *se* can be used to introduce an embedded infinitival interrogative, as in sentence (1) (Kayne 2005, p. 9):

(1) Gianni non sa se partire. (G neg knows if leave<sub>infin</sub>)

On the other hand, French disallows a controlled infinitive with *si*, as in sentence (2) (*ibidem*):

(2) \*Jean ne sait pas si partir. (J neg know not if leave<sub>infin</sub>)

Although it could be fully plausible to expect that this inter-linguistic difference might simply represent an isolated differential property (namely, that these two different possibilities do not correlate or cluster with any other grammatical property), according to Kayne's analysis the difference in the behaviour of *si* and *se* in their respective languages extends also to another linguistic property: «the relative placement of pronominal clitic and infinitive» (Kayne 2005, p. 10). In fact, those Romance languages which, like French, do not allow an infinitival interrogative to be introduced by *if* have the order clitic-infinitive, while languages like Italian have the order infinitive-clitic. Hence, this relatively small difference between two relatively close languages like Italian and French would be due to an individual parameter underlying two distinct clusters of syntactic properties.

### 3.3.1.1 – Kayne's argument for parametric theory

In addition to presenting a new approach to the investigation of morphosyntactic variation, Kayne proposes an interesting argument supporting both the microparametric approach and parametric theory in general. Since «work in microparametric syntax should [...] begin to give us some sense of a lower bound for the number of parameters (which in turn will bear on questions of learnability/acquisition)», Kayne is well aware that the comparison of closely related languages, by raising the question of how many distinct microparameters there are, may be subject to significant criticism if the minimum number of parameters required to account for all possible languages exceeded the amount assumed to be viable for language acquisition (Kayne 2000, p. 7). In other words, if the number of parametrically distinct languages/dialects were too large to be accounted for by a conceivably small number of parametric options, parametric theory would lose all its traditional appeal for representing an elegant solution to Plato's problem:

Under the assumption that acquisition proceeds by parameter setting, the child does not pick its language whole out of a set consisting of all possible languages. Rather, it sets individual (syntactic) parameters, the end result of which is (the syntactic component of) a grammar. If the number of possible languages were so large that the number of parameters the child had to set was unmanageable (i.e. not learnable in the amount of time available), there would indeed be a problem. (Kayne 2000, p. 8)

Considering the question of how many *possible* languages there are necessarily poses the question of how many *existing* languages there are. Starting from the general assumption that «there is no syntactically significant distinction to be drawn between “language” and “dialect” and no justification for neglecting the latter», according to Kayne the conservative – and still manageable – estimate suggesting that «the number of languages presently in existence is 4000-5000» turns out to be too optimistic (Kayne 2000, p. 7). The argument brought forward by Kayne specifically considers the high rate of syntactic variation in the territory of present-day Italy and the distinct varieties of English. With regard to Italian dialects, from a microparametric perspective the rate of syntactic variation is indeed particularly significant:

[...] in Northern Italy alone one can individuate at least 25 syntactically distinct languages/dialects solely by studying the syntax of subject clitics. More recently, I have had the privilege of participating in a Padua-based syntactic atlas/(micro)comparative syntax

project with Paola Benincà, Cecilia Poletto, and Laura Vanelli, on the basis of which it is evident that one can easily individuate at least 100 syntactically distinct languages/dialects in Northern Italy. A very conservative estimate would be that present-day Italy has at least 500 syntactically distinct languages/dialects. 500,000 would in consequence, I think, then be a very conservative extrapolation to the number of syntactically distinct languages/dialects in the world at present. (Kayne 2000, p. 7)

As far as the number of possible languages is concerned, «adding in those languages/dialects that have existed but no longer exist and those that will exist but do not yet exist» leads Kayne to estimate that «the number of syntactically distinct (potential) human languages is substantially greater than 5 billion» (Kayne 2000, p. 8). However, this does not represent a problem, as «the number of independent binary-valued syntactic parameters needed to allow for 5 billion syntactically distinct grammars is», from a strictly mathematical viewpoint, «only 33 (2 raised to the 33rd power is about 8.5 billion)» (*ibidem*). Starting from the assumption that parameters are features of functional elements, and that even if «the number of functional elements in syntax is not easy to estimate, [...] at the same time [...] 100 would be a low estimate» (Kayne 2005, p. 14), by postulating that the number of parameters which the learner is able to manage is on the order of a hundred, then the corresponding number of possible grammars characterizable by this set of binary options is «innocuously, over one million trillion trillion» (Kayne 2000, p. 8). This increase in the number of independent parameters would not represent a problem as well, since the child does not have to search through all the set of possible grammars but simply to set one parameter after the other until target grammar's final state is reached:

There is no problem here (except, perhaps, for those who think that linguists must study every possible language), since neither the language learner nor the linguist is obliged to work directly with the set of possible grammars. The learner needs only to be able to manage the task of setting the 100 parameters (or whatever the number is), and the linguist needs only to figure out what they are (and what the accompanying principles are, and why they are as they are). (Kayne 2005, p. 14)

In these terms, Kayne's argument represents an effort not only to reconcile the basic assumptions of microparametric syntax with the implications of the Borer-Chomsky conjecture, but also to strengthen both the descriptive and the explanatory power of parametric theory. In fact, although the Borer-Chomsky conjecture constitutes a constraint on the nature of parametric variation, the fact that each functional element is associated

with an independent parameter not only «is as constrained as minimalist syntax is» but, at the same time, it allows parametric variation to be far richer than originally thought (Rizzi 2013, p. 316). Therefore, from this point of view constraining the notion of possible parameter has limited the power of parametric theory to all appearances, while making the latter more suitable for being incorporated into Chomsky's MP.

### *3.3.1.2 – Microparametric syntax as an alternative to macroparameters*

Finally, while microparametric syntax could seem to simply represent a reaction to the crisis of the P&P model by means of revamping the traditional approach to parameter investigation, this perspective on linguistic variation also expresses serious reservations about the validity itself of the notion of macroparameter. On the one hand, the fact that microparametric syntax fully embraces the Borer-Chomsky conjecture means that the more traditional view that there can be parameters expressed on principles is sharply refused:

Now a widespread idea about syntactic parameters is that they are limited to being features/properties of functional elements, as opposed to ever being features of lexical elements. But since functional elements are part of the lexicon, then this limitation means that syntactic parameters are nonetheless necessary features, or properties, of elements of the lexicon. [...] Limiting syntactic parameters to features of functional heads is also intended to exclude the possibility that there could be a syntactic parameter that is a feature of no element of the lexicon at all – for example, there could presumably not be a parameter of the sort “language  $L_i$  has or does not have bottom-to-top derivations”. (Kayne 2005, p. 4)

On the other hand, the microparametric approach is based on a series of epistemological expectations that go beyond, if not against, the concept of (macro)parameter as encoding a wide range of linguistic variation. First, it is self evident that the comparison of closely-related languages, especially if performed under the assumption that the locus of parametric variation is exclusively identified with the functional categories, will tend to focus its attention also on aspects of cross-linguistic variation whose scope is far less broad than those clusters of properties traditionally dealt with by macroparameters. According to Kayne, to consider any instance of parametric variation «independently of the degree of “drama” or range of effects associated with any particular parameter» is a step forward that was really needed by parametric theory, since



the empirical value of the concept of parameter should not be confused with the degree of its effects on a certain grammatical system (Kayne 2005, p. 6). In fact, although the results of comparative syntax «can in some cases indeed take the form of a single parametric difference having a multiplicity of effects», in principle «it does not follow that every parameter, understood as a (simple) feature of some functional element, need have an equally wide range of effects» (Kayne 2005, p. 5):

It has occasionally been thought that the term “parameter” itself should only be used when there is such a notable, or dramatic, range of effects. I do not pursue that way of thinking here however, in part because what seems “dramatic” depends on expectations that may themselves be somewhat arbitrary. (*ibidem*)

Second, if the parametric differences resulting from the comparison of closely-related languages are assumed to have exactly the same epistemological value of those resulting from the comparison of more typologically distant languages, because of both their common “lexical” locus and the irrelevance of their superficial effects on grammar, a stronger implication of microparametric syntax is that macroparameters do not actually exist. Given the fact that «microparameters [...] are perfectly capable of participating in an explanation of a “cluster of properties”», according to Kayne «apparently macroparametric differences might all turn out to dissolve into arrays of microparametric ones» (Kayne 2005, p. 10). This idea could be expressed by the general conjecture:

Every parameter is a microparameter. (*ibidem*)

Accordingly, traditional macroparameters would turn out to be mere epiphenomena whose status has been somewhat overestimated because of the syntactic distance between the different grammars on which their formulation had been based. Kayne himself is somewhat vague about how the additive effects of some number of independent microparameters can actually act in concert to produce a macroparameter. However, his idea that «it may be that (some of) the clusters of syntactic properties that were under prominent discussion twenty-five years ago were too coarsely characterized», together with the expectation that «it may be that as research progresses a much finer-grained picture of syntax will substantially displace the one current twenty-five years ago» clearly counterposes the microparametric approach to the macroparametric one (Kayne 2005, p. 11).

### 3.3.2 – Baker's Parameter Hierarchy

The P&P assumption that there are likely to be structural relations between parameters and their settings in such a way that an implicational hierarchy among them can be posited has been always regarded as one of the most fundamental advantages of parametric theory for both accounting for the process of language acquisition and deriving robust typological generalizations. However, even before «a large portion of the field [...] moved away from looking for classical parameters to looking for microparameters, following the methodology and reasoning championed by Richard Kayne», generative grammar's efforts generally aimed at identifying the particular parameters distinguishing one language from another rather than deriving the relevant implicational relations among them (Baker 2008a, p. 1). In fact, as noted by Newmeyer, the only attempt at this kind of formulation was made only in the early 2000s, and it is represented by Mark Baker's book *The Atoms of Language* (2001) (cf. Newmeyer 2005, pp. 50-51).

The central assumption of Baker (2001) is that the typological generalizations occurring across languages can be explained by means of a hierarchy governing the implicational relations between the possible settings of each parameter. This idea, which is based on Chomsky's intuition that the fixing of the parameters of UG is «guided perhaps by a structure of preferences and implicational relations among the parameters of the core theory» (Chomsky 1993 [1981a], p. 7), according to Baker is empirically justified by the observation that some specific parametric settings can effectively exclude the effects of other parameters from occurring not only in a given grammar, but in every possible language:

Some parameters have a much greater impact on the form of an E-language than others do. Indeed, some parameters end up having no perceptible effect at all on the E-language, when some other parameter has been set in a way that makes the first one irrelevant in practice. (Baker 2001, p. 161)

Rather than considering these differences as a mere epiphenomenal effect due to the different interactions of each parameter with other elements of the grammatical system, Baker takes Chomsky's intuition literally and postulates the existence of «a purely logical order, in which parameters are ranked by their power to affect one another and their potentials for rendering each other irrelevant» (Baker 2001, p. 162). As an example of this kind of implicational relations between parameters and their settings, Baker considers

the Polysynthesis Parameter and the Head Directionality Parameter – namely, the very first parameters of his Parameter Hierarchy. Although these two parameters might seem to be independent of each other, as the former «determines whether the participants of an action have to be represented on the verb that expresses the action», whereas the latter «determines the order in which words are assembled into phrases», according to Baker's analysis the Polysynthesis Parameter and the Head Directionality Parameter are implicationally related (Baker 2001, p. 162). In fact, since a polysynthetic language like Mohawk allows both subject and object inflection markers on the verb and, accordingly, the corresponding full arguments are treated as adjuncts which can be attached to either side of the clause, in cases like this speaking of directionality becomes irrelevant, as «such noun phrases do not combine with the verb to form verb phrases, as they do in other languages» (*ibidem*):

The head directionality parameter is in practice irrelevant to the verb-object relationship in Mohawk. [...] Moreover, since the polysynthesis parameter applies to *all* the participants in an event, not just the direct object, similar effects are found in other kinds of phrases. If this reasoning is carried through consistently, the head directionality parameter could simply be irrelevant to Mohawk-style languages because the kinds of grammatical configurations it regulates never arise. (*ibidem*)

The fact that the Head Directionality Parameter – or, more precisely, all its potential settings – can actually be bypassed by one setting of the Polysynthesis Parameter means that the latter has a sort of logical priority over the former, thus constituting «a principled reason for ranking the polysynthesis parameter above the head directionality parameter» (Baker 2001, p. 163). Based on such conditions, this kind of reasoning can be extended to other parameters as well. At the same time, however, attention must be paid to the fact that not all parameters can be ranked with respect to each other; in fact, there is also the possibility that two parameters are logically independent – according to Baker, «probably because they characterize noninteracting aspects of language» (Baker 2001, p. 164). The structure of the implicational relations between parameters and their settings which emerges from the combination of these binary rankings can finally be made explicit by means of a comprehensive and systematic “table of languages”: a notational device similar to the periodic table of the chemical elements developed by Dmitri Mendeleev in the Nineteenth century.

Besides fulfilling the classical P&P program for typology, the greatest appeal of such a table of languages would actually lie in its predictive power. On the one hand, Baker's Parameter Hierarchy provides a well-grounded answer to the question of why certain typological features or clusters are more common than others across languages. Starting from the assumption that each parameter does in principle not imply any structural preference towards one of its settings, which would therefore be chosen by means of a process analogous to «the flip of a coin» (Baker 2001, p. 134), according to Baker the relative rarity of a language type is directly proportional to the number of parametric choices needed to characterize it. For example, as far as the relative frequency of head-initial versus head-final languages is concerned, the roughly equal occurrence of English-style word order and Japanese-style word order across languages would depend on the fact that the structural difference between these two language types depends only on a single binary parameter:

Since the difference between English-style word order and Japanese-style word order is attributable to a single parameter, there is only one decision to make by coin flip: heads, heads are initial; tails, heads are final. So we expect roughly equal numbers of English-type and Japanese-type languages. (*ibidem*)

Accordingly, the distributional difference between two distinct language types as those represented by SVO languages and the much rarer VSO languages would depend on the fact that the latter type requires two more parametric choices to be made than the former:

Within the head-initial languages, however, it requires two further decisions to get a verb-initial, Welsh-type language: Subjects must be added early *and* tense auxiliaries must host verbs. If either of these decisions is made in the opposite way, then subject-verb-object order will still emerge. If the decisions were made by coin flips, we would predict that about 25 percent of the head-initial languages would be of the Welsh type and 75 percent of the English type. This, too, is approximately correct. (*ibidem*)

On the other hand, although Baker's analysis was performed on the basis of strictly empirical data, that is, by observing the clusters of typological features occurring across a set of *existing* languages, the basic P&P assumption that typological variation ultimately depends on parameter setting interacting with universal principles makes his Parameter Hierarchy fully capable of predicting the combinations of properties characterizing the set

of *possible* languages. Since each parameter setting leaves open only a certain number of parametric possibilities, which in turn would be further restricted as more choices are made by the language learner, the typological facets assumable by any possible language would be necessarily limited to those that do not violate the implicational relations stated in the hierarchy:

Although we cannot predict the existence of new parameters in any detail (at least not yet), it is well within our grasp to predict the existence of new languages. Such languages should occur whenever the known parameters can combine in a way that has not been observed but that does not lead to a logical contradiction. It is reasonable to conjecture that languages with the theoretically possible combination of properties could be found if we looked for them. (Baker 2001, pp. 173-174).

Summing up, Baker's idea of Parameter Hierarchy is based on the assumption that UG is *overspecified* – namely, that the innate architecture of our language faculty determines *item by item* all the possible linguistic options selectable by the speakers. Accordingly, the acquisition task is reduced to choosing one of the possible grammars provided by UG, as this choice is dependent on the richness of UG's architecture and cannot violate the implicational relations between each parameter setting.

### 3.3.2.1 – Baker's argument for parametric theory

Besides constituting an unprecedented effort to find a typological application to the traditional concept of parameter hierarchy, the theory proposed in Baker (2001) also offers an argument for parametric theory. This argument, in turn, is based on the common thread in all of Baker's works that are reviewed in the present thesis: the epistemological value of linguistic typology, in both its underlying patterns and the most quantitative, statistical data deducible from it:

I should confess that not every linguist would assign so much importance to the notion of a parameter or understand it in exactly the same way. [...] Outside the [Chomskyan] paradigm, many linguists object (sometimes strenuously) to the terminology of parameters and some of the intellectual background associated with it, preferring a different terminology and different associations. But beneath the surface of controversy and debate, there is a growing understanding that the differences among languages are to be grouped into relatively stable patterns that do not arise as accidents of particular histories or cultures. (Baker 2001, p. 46).

As done by Kayne (2000), Baker's argument for parametric theory emphasizes the descriptive adequacy of a linguistic theory relying on the concept of parameter. However, while on the one hand Kayne's account aims at demonstrating that even a small number of parameters is capable of characterizing a great number of possible languages, on the other hand Baker's reasoning goes in the opposite direction. By strengthening the concept of parameter by means of an implicational hierarchy, which in turn sets narrow limits on possible grammars, Baker avoids such difficult issues as how many parameters can be feasibly handled by the language learner, or how many parametric options are needed to effectively account for the set of possible languages. Baker's idea is that, in accordance with the original P&P model as presented in Chomsky (1981a), not only does each parameter account for a certain cluster of typological properties, but that «the settings of different parameters» is somehow related as well, thus determining the existence of actual clusters of implicationally-related parameters (Baker 1996, p. 7). Accordingly, the issue arising from the growing number of syntactic parameters discovered is circumvented, as each binary setting overrides not only its opposite setting, but the whole set of implicationally related parametric choices governed in turn by the non selected setting. In fact, from this perspective language learning truly becomes more and more easy as the speaker's linguistic knowledge develops, since the class of possible grammars are progressively, and exponentially, restricted from a number (potentially) close to infinity to one.

As noted in the above quotation, Baker emphasizes the importance of typology in showing the existence of an actual limit to linguistic variation – a limit which would in turn depend on the interaction between different parameter settings. In this respect, although parametric theory could seem an oversimplified way to deal with linguistic diversity for its tendency «to see languages as falling into discrete types», a perspective based on the opposite assumption – namely, that variation is unconstrained – would not be able to account for those coherent patterns systematically emerging from the analysis of existing languages (Baker 2001, p. 82). This impossibility would in turn be clearly evident considering not only the fact that «languages that are close to the ideal types are much more common than languages that are far from them», but also that the possible range of linguistic variation does not exhaust all possibilities which would in principle be allowed by all the combinations of the generalizations characterizing each ideal type (*ibidem*):

[...] the real weakness of the "continuous variation" view is that it has a hard time explaining why many logically possible forms of language are extremely rare or nonexistent. However

you look at it, the range of possible human languages is much smaller than the range of conceivable ones. (Baker 2001, p. 83)

According to Baker, the descriptive inadequacy of any theory based on the continuous variation assumption is evident, for example, when considering what he refers to as the «two basic word orders», namely VO and OV (Baker 2001, p. 62). While speaking of two main word order types may seem a bit of a simplification, the fact that «these two word order patterns account for more than 95 percent of the languages of the world that care about word order at all» directly implies that «this kind of variation within tightly constrained boundaries is the sign of a parameter at work» (*ibidem*). In these terms, even the residual range of variation which is not covered by the currently available parameters could not justify embracing continuous variation. Rather, there is always the possibility «to include additional, finer-grained parameters to account for the “extra” languages» (Baker 2001, p. 84).

### *3.3.2.2 – The distinction between micro- and macroparameters*

Although there is no explicit reference to the concept of macroparameter, in Baker (2001) special emphasis is put on the distinction between those parameters having such a noticeable effect on grammar that they can define typologically distinct kinds of languages and those «additional parameters that happen to make only a relatively small difference in the qualities of the E-language as a whole» (Baker 2001, p. 125). On the one hand, a difference in a possible parametric setting has large repercussions «only if it is perfectly situated so as to maximize its interactions with other elements of the system» – which is basically the reason why certain typological features tend to co-occur in clusters (Baker 2001, p. 126). On the other hand, the effect of some parameters on grammar is not so evident, and therefore they behave similarly to «loose rocks that do not cause landslides because they are near to the bottom of a slope» (Baker 2001, p. 140):

Some simple changes in the recipe of a language have repercussions that affect the overall “feel” of the language in striking ways. [...] Other changes have the same logical character, but because of the way they happen to interact with other features of language their effects don’t spread through the language in the same way. These parameters do not create new types of languages but variations on one of the main types. (Baker 2001, p. 126)

An example of the more influential kind of parameters is the Head Directionality Parameter. As stated above, this parameter has massive superficial effects on language since it affects the construction of every phrase. On the other hand, the case of «a “minor” parameter» which «adds its influence to a “major” one» is represented by the Verb Attraction Parameter (or Verb Movement Parameter, as presented in Newmeyer (2005)), which can be formulated as follows (Baker 2001, p. 143):

(3) Verb Attraction Parameter:

Tense auxiliaries attract the verb to their position.

*or*

Verbs attract tense auxiliaries to their position.

(Baker 2001, p. 132)

According to Baker, the lower status of the Verb Attraction Parameter with respect to other more noticeable parameters depends on the fact that, considering two relatively similar languages as English and French, the different effects of its two possible settings are visible only in certain syntactic contexts. As shown by (4) and (5), when the auxiliary bears the tense inflection there is no discernible difference between English and French (Baker 2001, p.135):

(4) Jean a souvent embrassé Marie.

(5) John has often kissed Marie.

However, when there is no auxiliary and the main verb bears the tense inflection, in French the finite verb comes before the adverb, whereas in English it comes after, as in sentences (6) and (7) (*ibidem*):

(6) Jean *embrasse souvent* Marie.

(7) John *often kisses* Marie.

While the comparison between the Head Directionality Parameter and the Verb Attraction Parameter is surely striking, Baker's general distinction between *major* and *minor* parameters is grounded more on a descriptive perspective rather than on a strictly explanatory one. In fact, to say that «the effects of [...] the verb attraction parameter on E-language are not as far-reaching as the other parameters we have seen [the Polysynthesis



Parameter and the Head Directionality Parameter]» could seem a rather arbitrary conclusion, especially considering that, as Baker himself points out, sometimes even «a “minor” parameter like the verb attraction parameter makes its presence felt not only by positively influencing word order but by inhibiting other grammatical possibilities» (Baker 2001, p. 140). Moreover, the fact that every parameter ranking below the Polysynthesis Parameter can potentially be made irrelevant by one setting of the parameter immediately dominating it seems to suggest that there is no correlation between each parameter’s ranking and the visibility of its superficial effects on grammar – which thus depend exclusively on rather contingent, language-specific factors. The most notable example of this inconsistency is represented by the Null Subject Parameter. Although being traditionally regarded as the macroparameter *par excellence*, according to Baker’s hierarchy the Null Subject Parameter is actually at the very bottom of the list, even below the Verb Attraction Parameter:

As for the null subject parameter, this was originally presented as a matter internal to the Romance languages, distinguishing French (and English) from languages like Italian and Spanish. As such it would be relatively low in the table. Indeed, there are conjectures in the field that only a proper subset of the verb-attracting languages can be null subject languages in the original sense. If these conjectures are correct, then the null subject parameter would be ranked below the head attraction parameter and probably below its partner, the subject placement parameter, as well. That is where I put it. (Baker 2001, p. 168)

Summing up, while the parameters examined in Baker (2001) are all supposed to obey a strict logical ranking, their division into minor and major parameters does not depend on neither their position in the hierarchy nor other structural conditions, but only on the extension of the cluster of properties to which they can be related. In these terms, Baker’s notion of “major parameter” corresponds to the traditional definition of macroparameter: a binary option which alone is responsible of a cluster of typological proprieties.

### 3.3.2.3 – A revaluation of the macroparametric approach

More recently, the well-established distinction between macro- and micro-parameters has been complemented with a further dichotomy which does not refer to the extent of the parameter’s impact on language but to the locus of parametric variation. The

first of these categories, which were first introduced in Baker (2008a), is that of *lexical parameters*. As the name suggests, lexical parameters are those parameters which are consistent with the Borer-Chomsky conjecture, and which are therefore located in the (functional) lexicon. The second category is that of *grammatical parameters*, and it refers to those parameters whose locus, according to Baker, cannot simply be located in the functional lexicon as they are specified by UG itself:

There are some parameters within the statements of the general principles that shape natural language syntax. (Baker 2008a, p. 3)

While such a distinction does not negate the validity of either the concept of microparameter nor the microparametric approach to the study language variation, the implications of the notion of grammatical parameter run diametrically counter to the widespread assumption that «syntactic parameters [...] are limited to being features/properties of functional elements» (Kayne 2005, p. 4). Starting from the assumption that the Borer-Chomsky conjecture is actually «a hypothesis to be proven empirically, not an established result», and which could therefore concern not all existing parameters but only a limited number of them, Baker postulates the possibility that there can be parametric variation both inside and outside the lexicon (Baker 2008a, p. 3):

I, for one, am interested in the possibility that the Borer-Chomsky Conjecture may be false. The contrary view is that there can be variation in the grammar proper in addition to variation that can properly be attributed to the properties of particular lexical items. (*ibidem*)

As pointed out by Baker, «this macroparametric view [...] is in fact no different from the classical early 1980s view of the parameter», that is, from the idea that principles can be parameterized (*ibidem*). As such, Baker is admittedly aware of the fact that, on the one hand, while the concept of grammatical parameter is not implausible considering a minimalist perspective, the idea that the locus of all possible parametric variation does necessarily correspond to the lexicon is far more attractive than the idea that parameters are expressed on principles. However, when considering the broad typological differences triggered by the possible settings of such parameters as the ones governing Polysynthesis and Head Directionality, according to Baker there seems to be at work something deeper than some individual lexical elements working in concert:

[...] there seems to be a certain global unity to a head-final language as opposed to a head-initial language, or to a polysynthetic language as contrasted with a more isolating language, which seems more pervasive than can be attributed to any particular lexical item, or even to a small class of lexical items. Nor does the Minimalist Program necessarily rule out the possibility of classical/macroparameters in the sense of [grammatical parameters]. (Baker 2008a, p. 4)

According to Baker, the fact that the universal validity of the Borer-Chomsky conjecture has been taken for granted throughout the MP has led the majority of the linguists interested in parametric issues to adopt a microparametric approach based only on «the methodology of comparing closely related languages, dialects, and varieties» (Baker 1996, p. 8). However, while this particular approach is surely appropriate for discovering lexical parameters, on the other hand it cannot be effective at identifying any of what Baker defines as grammatical parameters, as the comparison is made only between languages which do not substantially differ in their typological properties:

Simply finding many differences among languages that are consistent with [the notion of lexical parameter] does not at all prove that [the Borer-Chomsky Conjecture] is true, because [the Borer-Chomsky Conjecture] is a universal statement. That is particularly so if much of the field is pursuing a methodology that presupposes that [the Borer-Chomsky Conjecture] is true and is designed to find parameters of this type, but which is not at all likely to find parameters of other kinds – as I claim is the case. (Baker 2008a, p. 3)

Similarly to what previously argued in Baker (2001), the argument proposed here for the revaluation of the macroparametric approach relies on a typological perspective. Leaving aside the possibility that all syntactic variation is macroparametric variation of the kind envisioned in the definition of grammatical parameters, as «for any binary-valued parameter one would expect to find two sharply different parametric clusters, with essentially no intermediate cases» (Baker 2008a, p. 8), according to the same reasoning it would equally wrong to suppose that, as predicted by the Borer-Chomsky Conjecture, «there are only microparameters and they are logically independent of each other» (Baker 2008a, p. 9). In fact, if the latter hypothesis were true, according to Baker the aspect of typological variation across languages would be way more scattered than it appears when examining the pervasive typological differences generated, for example, by the different settings of such parameters as the Polysynthesis Parameter and the Head Directionality Parameter:

If it is correct to reduce all macroparameters to a series of relatively independent microparameters in this way, then one would expect to find a relatively smooth continuum of languages. Along the polysynthesis dimension, languages would range from those that have few or no dislocation constructions, to those that have several, to those that have many, to those that have a complete set. Similarly, one could have a wide variety of languages when it came to head-complement order, with pure head-final languages like Japanese and pure head initial languages like English simply being the special cases in which all of the categories happen to be set for the same complement-head order. (Baker 2008a, p. 9)

All things considered, Baker indicates the coexistence of macroparameters and microparameters as the only viable scenario, as «the macroparametric-plus-microparametric approach predicts that there will be more languages that look like pure or almost pure instances of the extreme types, and fewer that are roughly equal mixtures» (Baker 2008a, p. 10).

In order to further demonstrate both the plausibility and feasibility of a macroparametric approach in a minimalist context, Baker (2008a,b) proposes two new parameters which seem to be not compatible with the Borer-Chomsky Conjecture. These grammatical parameters are formulated as follows:

- (8) a. The Direction of Agreement Parameter:

[A functional head] F agrees with DP/NP only if DP/NP asymmetrically c-commands F.

- b. The Case-Dependency of Agreement Parameter:

[A functional head] F agrees with DP/NP only if F values the case feature of DP/NP or vice versa.

(Baker 2008b, p. 155)

According to Baker, the Niger Congo (NC) languages behave differently from the Indo-European (IE) languages with respect to the agreement of all functional heads. In this respect, the positive setting of the Direction of Agreement parameter and the Case-Dependency of Agreement parameter is mutually exclusive: while the former characterizes the NC languages by predicting that there is an asymmetrical c-command relation between the agreed with NP and the agreeing head, the latter correctly predicts that agreement in IE languages depends on Case. In the following examples, Baker discusses the application of these parameters to subject agreement on the finite Tense node: sentence (9) comes from the Bantu language Kinande, while sentence (10) comes from Yiddish. In both (9) and (10), an argument other than the thematic subject has been displaced to Spec, TP. Although both Kinande and Yiddish normally show agreement between the finite

verb and the preverbal subject in simple clauses, «when something other than the thematic subject moves to Spec, TP [...] the Bantu verb agrees with the phrase that has moved to Spec, TP», which in (9) corresponds to the object, while on the other hand «the IE verb agrees with the nominative NP (the thematic subject) regardless of where it appears in the structure», as shown in (10) (Baker 2008a, pp. 12-13):

- (9) Olukwi si-lu-li-seny-a bakali (omo-mbasa). (Diesing 1990)  
 wood.11 NEG-11S-PRES-chop-FV women.2 LOC.18-axe.9  
 “WOMEN do not chop wood (with an axe).”
- (10) ...az vayn ken men makhn fun troybn oykh.  
 that wine can one make from grapes also  
 “(You should know)...that one can make wine from grapes also.”

In both Kinande and Yiddish, this particular behaviour of Tense extends uniformly to the other functional heads, thus being «a relatively general feature of the language, not one that is tied to a particular head or construction» (Baker 2008b, p. 155). In order to exclude the possibility that, even in this case, variation can be determined by a series of lexical parameters, a statistical test is proposed: if the difference between the Niger Congo languages and the Indo-European languages were of microparametric nature, according to Baker there should be many mixed languages of different kinds «in which roughly half the functional heads show the IE behavior and the other half show the NC behavior»; on the other hand, a macroparametric account would expect to find many languages in which all functional heads behave consistently, according to either the Direction of Agreement parameter or the Case-Dependency of Agreement parameter, and «only a few cases that are intermediate or hard to classify» (Baker 2008a, p. 17). The analysis of about 66 languages out of the 100 languages in the core sample of *The World Atlas of Language Structures* does indeed confirm the macroparametric hypothesis, as all languages examined obey either the Direction of Agreement parameter or the Case-Dependency of Agreement parameter.

Finally, as Baker points out, probably a microparametric approach to language variation would have never discovered two parameters like (8a) and (8b) simply by comparing a set of closely related languages.

### 3.4 – Against a parametric approach to typological variation

While in the previous section two main examples have been given of what Baker means when he notes that, «within the Chomskyan paradigm, there are many linguists who accept the terminology of parameters but have somewhat different views about exactly what a parameter is and what the best examples are» (Baker 2001, p. 46), in this section special attention will be paid to some of the major criticisms to the view that cross-linguistic grammatical variation is encoded by syntactic parameters. This position has been criticized most notably by Frederick J. Newmeyer (2004, 2005). Although some of these works' arguments against both the macroparametric and the microparametric approach have lately been proposed in other works such as Boeckx (2008, 2010), here I will focus my attention only on Newmeyer's claims, as they do not aim at overcoming the shortcomings of current parametric theory but, more radically, at undermining its theoretical foundations.

#### *3.4.1 – Against a generative account for typological generalizations*

Before looking at Newmeyer's «frontal assault [...] on the parametric approach to grammar», it is well worth looking at one of the most important, if not the most important, reasons why, according to Newmeyer (2005), language typology alone cannot be an argument for determining the range of possible human languages (Newmeyer 2005, p. 73). In explaining this view, it is no accident that it is precisely Baker's statistical approach that is called into question. Although Baker's arguments for the idea of Parameter Hierarchy and the concept of (macro)parameter are based on the premise that the range of the observed cross-linguistic variation delimits what is possible or impossible in language, thus defining what all possible grammars necessarily have to look like, Newmeyer claims that «there is simply no motivation for declaring that a grammatical feature is 'impossible' if it happens to be missing from some arbitrary number (or percentage) of languages» (Newmeyer 2005, p. 4). In fact, it should be clear not only that a limited sample of languages, numerous though this sample may be, is not enough to make a universal claim, but also that taking for granted the validity of a P&P approach does in principle exclude to identify the influence of any other factor on language variation. In this respect, according to Newmeyer:

[...] the problem is that such a claim of impossibility is so theory-laden that there is no way to evaluate it independently of the complex web of hypotheses that characterize early twenty-first century principles-and-parameters syntax – hypotheses that themselves are constantly undergoing revision. (Newmeyer 2005, p. 3)

Moreover, another possible fault attributable to a parameter-setting approach is that there is no guarantee that, given the fact that some grammatical features have either been observed or not in any language, these features «are *necessarily* universally occurring or universally prohibited as a consequence of our biological preprogramming (for grammar per se or for broader faculties not specific to language)» (Newmeyer 2005, p. 6). To this end, Newmeyer does not exclude that the seemingly logical impossibility of some non-existing grammars can actually be accounted for without resorting to the notion of parameter. Instead, he argues that the fact that some structures are *incidentally* prohibited may be attributed to functional factors regarding, for example, communication efficiency «or other aspects of language use» (*ibidem*).

### 3.4.2 – Newmeyer's critique of parameters

Rather than reviewing a series of specific syntactic parameters, Newmeyer's critique of parameters consists essentially in comparing the putative theoretical advantages of a parametric approach to language variation over a rule-based approach. By means of minimizing the former with respect to the latter, Newmeyer argues not only that the P&P model shows many deficiencies in itself, but also – and more importantly – that these deficiencies do not affect the efficiency of a rule-based model accounting for typological variation.

Newmeyer begins by listing some the central features of the P&P model and the rule-based model proposed as its alternative. The former model, which is described as the «currently predominant linguistic theory», according to Newmeyer is based on the following assumptions (Newmeyer 2005, p. 73):

- (11) a. Principles of Universal Grammar (or, more recently, a set of functional projections provided by UG), which have
- b. Different parameter settings for different languages (thereby accounting for language-particular differences).
- c. By means of [11a] and [11b], typological variation is accounted for.

- d. A residue of marked (language-particular) morphosyntactic properties.  
(Newmeyer 2005, p. 73)

On the other hand, the alternative position represented by the latter model is characterized as follows:

- (12) a. Unparameterized principles of Universal Grammar.
  - b. Language-particular rules constrained by these UG principles.
  - c. Extragrammatical principles that account for typological variation.
- (*ibidem*)

Although the P&P model does not require any further explanation, it is important to point out that the “language-particular rules” included in Newmeyer’s proposal are not simply meant as mere formal devices whose aim is to explicitly describe those linguistic phenomena whose explanation does not seem to obey any principled explanation. Rather, by the term “language-particular rules” Newmeyer means «parameter-settings ‘detached’ from the parameters themselves (which are hypothesized not to exist)» (Newmeyer 2005, p. 74). What Newmeyer actually aims at doing by proposing such an independence between parametric options and syntactic parameters is, as it will be discussed soon, to equalize the concept of parameter to that of rule, thus undercutting the appeal of the P&P model.

As far as the evaluation of the alleged advantages of a parametric approach is concerned, the first aspect which is regarded as a proof of the superiority of the notion of parameter over that of rule is the *descriptive simplicity* of the former over the latter. According to Newmeyer, however, there is actually no point in taking this assumption for granted. What Newmeyer argues here is very simple: «if as many parameter settings are needed in a parameter-based model as rules would be needed in a rule-based one and the former turn out to be as complex as the latter, then clearly nothing is gained by opting for parameters» (Newmeyer 2005, p. 77). Taking the Head Directionality parameter as an example, since the two parametric options entailed by this parameter’s binary setting can ideally be handled by two language-particular rules – namely, «complements are to the right of the head» and «complements are to the left of the head» – there is no point in considering a parametric account of this specific kind of typological difference as descriptively simpler than a rule-based approach (Newmeyer 2005, p. 74).



The second aspect of parametric theory that is debated in Newmeyer (2004, 2005) is *binarity*, an idea which is inapplicable to rules in general. In this regard, Newmeyer claims not only that «there is little evident binarity in morphosyntax», but also that many parameters cannot be assumed to be binary at all (Newmeyer 2005, p. 79):

As far as parameters are concerned, some have indeed been assumed to have binary settings, such as those that determine whether a language is configurational or not, whether it has (overt) *wh*-fronting or not, and so on. But many others are not (evidently) binary. (Newmeyer 2005, p. 80)

Among the examples proposed, the most relevant one is represented by the governing category parameter proposed by Manzini & Wexler (1987). According to Newmeyer, when looking at this parameter «the possible binding domains across languages have been argued to be in a subset relation with respect to each other, rather than contrasting in a binary fashion» (*ibidem*).

The next claim made by Newmeyer regards the fact that, while it is commonly argued that the number of syntactic parameters needs to be reasonably small to be viable for language acquisition, not only does this prerequisite not correspond to what has been proposed by parametricians, but it also is actually unrealistic. As it regards the number of syntactic parameters proposed within the P&P model, Newmeyer points out that «certainly hundreds have been proposed since the notion was introduced around 1980», thus running counter that same *smallness of number* identified as an advantage of parametric theory over language-particular rules (Newmeyer 2005, p. 81). Moreover, although there are surely some general parameters whose different settings characterize very broad linguistic types, according to Newmeyer «others have the appearance of being uncomfortably language-particular» (*ibidem*). Regarding the number of parameters which could be plausibly specified by UG, according to Newmeyer the estimate proposed in Kayne (2000) on the basis of the empirical data provided by syntactic microvariation is too optimistic for accounting for linguistic variation. Newmeyer's main argument is that, although Kayne calculated the minimum number of parameters needed to characterize the set of all parametrically distinct languages/dialects, he actually had to consider the set of all possible languages instead:

[...] in order to account for microparameters and the 'some number substantially greater than five billion' grammars that might exist in the world (Kayne 2000b: 8 [=2000, A.R.]), Kayne

calculates that only thirty-three binary-valued parameters would be needed. His math may be right, but from that fact it does not follow that only thirty-three parameters would be needed to capture all of the microvariation that one finds in the world's languages and dialects. In principle, the goal of a parametric approach is to capture the set of *possible* human languages, not the set (however large) of actually existing ones. (Newmeyer 2005, p. 83).

Speculating that «the number of such languages is in the trillions or quadrillions», Newmeyer suggests that the number of parameters needed to handle such a range of linguistic variation may actually be «in the thousands (or, worse, millions)» (*ibidem*). Based on such conditions, Newmeyer rejects both the viability of the process of parameter setting and the advantage of a parameter-setting model over a rule-based model. On the one hand, starting from the assumption that the number of parameters is much larger than Kayne expected, it would be highly implausible, if not absurd, to suggest not only that the child should set something like a thousand of parameters, but also that language faculty would have been endowed by human evolution «in such an exuberant fashion» (*ibidem*).

Fourthly, in his treatise against parametric theory, Newmeyer criticizes the idea that parameters are *hierarchically/implicationally organized*. In this regard, the target of criticism is represented by the Parameter Hierarchy proposed in Baker (2001). Leaving aside some purely empirical difficulties found in Baker's typology, the main aspect discussed by Newmeyer is «the idea that the rarity of a language type is positively correlated with the number of 'decisions' (i.e. parametric choices) that a language learner has to make» (Newmeyer 2005, p. 85). Although the Parameter Hierarchy predicts, for example, that the number of non-polysynthetic languages should be roughly equal to the number of polysynthetic ones as these two types are separated by a single parametric choice, according to Newmeyer this statement is false because the former are many more than the latter. Along the same lines, another example which should falsify Baker's idea is that, while the Null Subject Parameter is at the very bottom of the hierarchy, thus «implying that null subject languages should be rarer than verb-initial languages», null subject languages do actually represent «a solid majority of the world's languages» (*ibidem*).

Another aspect related to the implicational relations between parameter settings is the claim that parametric theory allows «the prediction of (unexpected) clusterings of morphosyntactic properties» (Newmeyer 2005, p. 76). Here Newmeyer argues that the cluster of properties which were initially associated with the Null subject parameter has

been shown not to hold. The earlier account of this macroparameter, as originally formulated in Rizzi (1982), stated that the possibility of null thematic subjects in tensed clauses, null non-thematic subjects, subject inversion and the violation of *that*-trace filter would be implicationally related, thus predicting the following clusters of features (Newmeyer 2005, p. 89):

(13)	NULL TS	NULL NTS	SI	THAT-T
	yes	yes	yes	yes
	no	yes	yes	yes
	no	no	no	no

However, Newmeyer points out that «still other language types exist, or at least appear to» (*ibidem*). For example, Brazilian Portuguese and Chinese allow null subject, but not subject inversion. Moreover, even a further fragmentation of the aforementioned clusters taking into account these two intermediate types, as the one postulated by Safir (1985), did not stand the test of time as more languages were added to the comparison – most notably, by Gilligan (1987). Given the fact that it has showed no evidence for the existence of the tight parametric cluster which was at the basis of its formulation, according to Newmeyer the Null Subject parameter is a mere epiphenomenon, as it shows no consistency in a cross-linguistic dimension.

Nevertheless, Newmeyer's criticism does not stop at traditional macroparameters. In fact, another aspect criticized in Newmeyer (2005) regards the parametric approach to language variation carried out in Minimalism. On the one hand, according to Newmeyer, the shift from the traditional idea that principles can be parameterized to the minimalist approach, which regards parameters as features on functional heads in the lexicon, «makes it all but impossible to predict any significant degree of clustering» as these *lexical parameters* are less pervasive in their range than the earlier, *grammatical* ones (Newmeyer 2005, p. 94). On the other hand, while the Lexical Parameterization Hypothesis certainly has its advantages in locating parametric variation in the lexicon, Newmeyer points out that «in no framework ever proposed by Chomsky has the lexicon been as important» but, at the same time, «have the properties of the lexicon been as poorly investigated» as it is in Minimalism (Newmeyer 2005, p. 95, n. 9):

Puzzlingly from my point of view, the relocation of the site of parametric variation from grammars of entire languages to functional heads is often portrayed as a major step forward.

[...] But the price paid for this 'radical simplification' is both an explosion in the number of functional categories needed to be posited within UG and, more seriously, the transfer of the burden for accounting for language-particular differences from properties of UG per se to idiosyncratic properties of lexical entries in particular languages. (Newmeyer 2005, p. 95)

In these terms, Newmeyer's conclusion is that both macroparameters and microparameters have failed at fulfilling the original promise of accounting for typological variation by means of a finite set of binary settings. While the traditional macroparametric approach was overly ambitious in trying to correlate unexpected clusters of typological properties, the more modern microparametric approach has not enjoyed a better fate as, «given the LPH on the other hand, parameter settings seem to differ little in their degree of generality from language-particular rules» (Newmeyer 2005, p. 98).

Finally, Newmeyer argues that there is no point in assuming that UG's architecture is overspecified – namely, that while language faculty leaves a certain range of options open, it also provides a discrete series of (parametric) options. Although to say that UG allows syntactic variation only by leaving certain possibilities underspecified means that the central P&P assumption that «parameters and the set of their possible settings are innate (and therefore universal)» is no more valid (Newmeyer 2005, p. 76), Newmeyer points out that this position «does not in and of itself entail a rejection of innate UG principles per se» (Newmeyer 2005, p. 98). Accordingly, an underspecified UG composed only by language universals and assisted by a set of language-specific rules would still be capable of addressing the issue of the poverty of the stimulus, as the learner could still rely on a set of innate conditions limiting the number of possible grammars.

In conclusion, regardless of the specific arguments against the individual advantages of the P&P model for language variation, what makes Newmeyer's criticism of the parametric approach to grammar particularly appropriate is the observation that, once the fragmentation of parameters reaches the point that parameters themselves cannot be distinguished from language-particular rules, «there is no increase in descriptive elegance, economy, or whatever» over a rule-based model (Newmeyer 2005, p. 78). In these terms, Newmeyer's claim should not be regarded as a sterile and negative criticism, but as a cause for reflection which could be useful to further develop and strengthen the concept of parameter in Generative Grammar.

### 3.5 – Some perspectives on the debate about parametric theory

Although the parallel development of the macroparametric and the microparametric approach to linguistic variation attests to the continuing validity of parametric theory, some differences between how the notion of parameter has been conceptualized in these two approaches are nonetheless irreconcilable. Moreover, even within each of these two approaches there can be found some particular aspects which, while being perfectly coherent with their respective parametric perspective, upon closer examination appear to be either somewhat vague or, at worst, clearly contradictory with the original P&P model's intent of resolving the inherent tension between descriptive and explanatory adequacy. As a result, not only does Generative Grammar seem to be forced to choose between two mutually exclusive conceptions of parameter, but this division does also make parametric theory vulnerable to some of the criticisms which have been made against it. With this in mind, in this section some perspectives on the present day's parametric theory will be outlined in order to show how the P&P model, if considered in its unity, can not only find some strong points of convergence between the macroparametric and the microparametric approach to linguistic variation, but also effectively respond to the main criticisms made against the validity of its working hypotheses.

#### *3.5.1 – The crucial difference between the conceptions of parameter in the microparametric and the macroparametric approach and their relation to explanatory adequacy*

Comparing Kayne's microparametric approach with Baker's traditional macroparametric approach reveals not only a difference in their perspectives on language variation, but also a sharp discrepancy in their respective ways of conceptualizing the very notion of parameter. As pointed out above, according to Kayne's view the traditional concept of macroparameter has very little, if not none, epistemological value as microparameters are perfectly able to account for both individual and multiple syntactic properties. On the other hand, Baker regards microparameters alone as insufficient to handle parametric variation for the very reason that «in many cases [...] the "cluster" of properties accounted for by each [micro]parameter includes only one member» (Baker 1996, p. 7). In these sense it could be argued that, while Kayne's conception of parameter

is independent from the number of syntactic properties generated by its possible settings, Baker's idea is that parameters in the strictest sense of the word do necessarily have to account for a combination of multiple syntactic properties. Contrary to Baker's claim, Kayne's analysis clearly entertains the possibility that a single parameter might simply account for an isolated property distinguishing one grammar from another. In fact, the ideal goal of microparametric syntax is exactly that of approximating the results of a controlled experiment by isolating the effects of individual parameters on two closely-related languages as much as possible – namely, by progressively shrinking each parametric cluster until the point that it assumes the shape of an individual syntactic property covarying between two languages:

To the extent that one can find languages that are syntactically extremely similar to one another, yet clearly distinguishable and readily examinable, one can hope to reach a point such that the number of observable differences is so small that one can virtually see one property covarying with another. (Kayne 2000, pp. 5-6)

However, such an “impoverished” concept of parameter does not entail any contribution to explanatory adequacy, as it is merely a means of describing an individual difference between two grammars without making any contribution to the development of a theory of language acquisition – that is, without justifying this epiphenomenal difference on grounds of its relation to those principles of UG which provide an answer to how the child develops his own linguistic competence. Although, on the one hand, it is true that an isolated differential property clustering with no other intra-linguistic difference might plausibly be generated by a syntactic parameter, on the other hand there is no particularly principled reason warranting that this is actually the case. Baker seems well aware of this fact when he claims that these specific microparameters «are not parameters in the original sense at all, in spite of the fact that the terminology is retained» as they provide at best only a descriptively adequate account of the particular linguistic phenomena they are responsible for (Baker 1996, p. 7):

To the extent that this happens, the very idea of a parametric cluster is called into question, and microparameters become no more general than the constructions of traditional grammar. [...] Thus, although the P&P notion of a principle has proved very fruitful in many different settings, the notion of a parameter has not fulfilled its original promise. (*ibidem*)

Interestingly enough, this issue is exactly the same one which would be addressed nine years later by Newmeyer when he argued that «the problem with such an account is that the word ‘parameter’ is used as nothing more than a synonym for the word ‘rule’» (Newmeyer 2005, p. 78). In addition to the claim that many alleged parameters are so language-particular that nothing would be gained by preferring them over language-specific rules, according to Newmeyer the parallel coexistence of very weak parameters empirically equivalent to rules and stronger, more abstract parameters in the same linguistic theory would force linguists to make a choice in favor of either a P&P model integrated with a set of language specific rules or, as Newmeyer advocates, a fully coherent rule-based model. Therefore, the results of this choice would not be favorable to the former way of conceptualizing linguistic variation, and this exactly for the sake of epistemological simplicity:

Such a treatment undercuts the attractiveness of a parametric approach. That is, it is no longer a matter of comparing a theory with parameters (and all their virtues) with a theory with rules (and their lack of virtues). Rather, it is a matter of comparing a theory with parameter settings *and* rules versus one with rules alone. On the basis of Occam’s razor, one would be forced to renounce the idea of any *a priori* desirability of a parametric theory. The dilemma here is that if the scope of parameters is expanded [...], then ‘parameter’ has simply become a synonym for ‘rule’ [...]. If parameters are kept simple and few in number, then grammatical theory needs to characterize the properties of rules, just as it would have to in a purely rule-based approach. (Newmeyer 2005, pp. 78-79)

However, one issue that should not be overlooked is that the superiority of the notion of parameter over that of rule, and hence the success of the P&P model in Generative Grammar, lies not simply in the fact that small changes in the parameters would lead to explicit effects on generated grammatical structures, but that each of these small changes would determine strong correlations between two or more otherwise independent linguistic phenomena which, in some cases, are undetermined by evidence available to the language learner. As noted in Roberts & Holmberg (2010):

Now, if each parameter value determines a cluster of disparate syntactic features, then explanatory adequacy is enhanced, especially if certain features are readily accessible to acquirers on the basis of impoverished evidence while others are hardly likely to be easily accessible. In this case, arriving at a parameter value determining both the accessible and relatively inaccessible feature gives us a simple account of how the inaccessible feature can be acquired, thus accounting for an aspect of the poverty of the stimulus to language

acquisition and thereby, again, reaching explanatory adequacy. At the same time, other things being equal, a 'typological' prediction is made: the inaccessible feature will be acquired whenever the acquired one is, since both reflect the same abstract property of Universal Grammar, the setting of a given parameter to a given value. (Roberts & Holmberg 2010, p. 15)

Based on such conditions, the criticisms made in Newmeyer (2005) against the advantage of a parametric approach over a rule-based model are well-grounded provided that the status of parameter be assigned to those binary settings which account for a minimal pair of syntactic properties – that is, on condition that the ideal goal of Kayne's microparametric syntax is taken literally. In this scenario, each setting would indeed be indistinguishable from a language-particular rule. At the same time, however, Newmeyer's argument loses all its appeal when faced with the fact that a rule-based account to language variation represents an explicit renunciation of explanatory adequacy compared to the original, "cluster-based" conception of parameter as defended by Baker's macroparametric perspective.

Summing up, while «parameters at their finest-grained» provide the syntactician with an extremely powerful lens by which to analyze and describe the minimal units of syntactic variation (Kayne 2000, p. 6), stating that a parameter can potentially determine only one individual syntactic property instead of (even a) minimal cluster of them represents a retreat from explanatory adequacy as, in the former case, there would be no means to demonstrate that the alleged parameter in question entails an explanatorily adequate account of its effect on target language or dialect. Accordingly, stating that every parameter has complex effects across a range of surface phenomena represents, contrary to Newmeyer's claims, a crucial advantage over a rule-based approach, and this precisely because «parameters allow us to collapse cross-linguistic differences into single abstract properties of grammars, while language-specific rules are just that: language-specific» (Roberts & Holmberg 2005, pp. 541-542).

### *3.5.2 – Macro- vs micro-parameters: a terminological and epistemological issue*

In addition to providing an answer to Newmeyer's main argument against the P&P model, the idea that parameters in the strictest sense of the word have to account for a cluster with two or more ancillary morphosyntactic phenomena potentially allows



parametric theory to dispose of the terminological opposition between macro- and micro-parameters. If, on the one hand, the distinction between the macroparametric and the microparametric approach to language variation is still desirable, inasmuch as it conveys the fact that syntactic variation can be investigated from two main perspectives – that is, by either looking at samples of unrelated languages or comparing languages and dialects which are very closely related – on the other hand talking about a difference between the notion of macroparameter and that of microparameter becomes not only unnecessary, but also potentially misleading.

Although the contrary is often declared, the distinction between macro- and micro-parameters has sometimes been improperly related to a presumed difference not only in the scope of their respective syntactic effects, but also in the locus of their variation. In this regard, an example of the significant relation between the notions of macro- and micro-parameter and those of grammatical and lexical parameter can be found in Baker's approach to language variation. On the one hand, Baker (2008a) starts from the well-grounded assumption that the locus of parametric variation cannot in principle be related to the fact that a syntactic parameter may have either a small or a substantial grammatical effect on a grammar:

Calling [the notion of grammatical parameter] a macroparametric view is something of a misnomer. It is perfectly possible that a lexical parameter consistent with [the Borer-Chomsky Conjecture] could have a substantial impact on the language generated, particularly if it concerned some very prominent item (such as the finite Tense node). Conversely, it is quite possible that a syntactic parameter of the sort envisioned in [the notion of grammatical parameter] could have only a small and hard-to-notice impact on the language generated. (Baker 2008a, p. 4)

This idea is indeed consistent with the view that, as already pointed out by Kayne (2005), «first, what counts as a big and impressive difference as opposed to a small localized difference is a judgment call, impossible to make precise» and secondly, as previously noted in Baker (2001), «presumably whether a parameter (either lexical or grammatical) has large-scale effects on language has more to do with accidents of frequency and with how different factors happen to interact than with the inherent nature of the parameters themselves» (Baker 2008a, p. 5):

Even more importantly, there is no reason to think that the causes of relatively big, impressive differences among languages are any different in kind from the causes of

smaller, easily overlooked differences. [...] To the scientist, the interesting question is not usually how common something is, but rather how revealing it is about the essence of things. (Baker 2008a, p. 5)

On the other hand, however, Baker claims that «calling [the notion of grammatical parameter] a macroparametric view [...] is not a total misnomer» (Baker 2008a, p. 4) and this because, although acknowledging that «the extent of variation issue by itself is hard to pin down, and is ultimately not so important for the science of linguistics», at the same time he does not fully embrace the idea that «the extent of variation question is *completely* irrelevant in practice» (Baker 2008a, p. 5). This subtle but still tangible connection between the notion of macroparameter, which should be solely meant as the discrete syntactic variation setting apart two typological distant grammars, and that of grammatical parameter derives directly from the idea that the «lexical parameters consistent with the Borer-Chomsky Conjecture will *tend* to have smaller scale effects than grammatical parameters», while «in contrast, the grammatical parameters [within the statements of the general principles that shape natural language syntax] are not intrinsically limited in this way» (*ibidem*). In fact, as far as the latter are concerned, according to Baker's view:

They might pertain to all headed phrases, or to all theta-role assignment relationships, or to all functional heads of a certain type, and hence transcend a small number of constructions. Large-impact parameters are not inherently interesting just because they are large-impact. But they might be heuristically significant because they tend to point to loci of variation in the grammar as opposed to the lexicon. (*ibidem*)

What Baker seems to imply here is that lexical parameters, being simply features individually assigned to each functional element, are in principle prone to determine only those small-scale differences which in turn are supposed to distinguish only closely related languages or dialects – namely, those differences which in Kayne (2000, 2005) were identified with the notion of microparameter. By the same reasoning, according to Baker the larger-scale cross-linguistic variation commonly observed when adopting a macroparametric perspective cannot possibly be attributed to the cumulative effect of numerous microvariants of the previous kind, but rather represents the influence of parameters whose locus of variation is not located in the functional lexicon, but in the architecture of UG itself. What follows from this assumption is that in this way, paraphrasing Newmeyer (2005), the word “macroparameter” is used as nothing more than

a synonym for the word “grammatical parameter”, and the same happens for the notion of microparameter and that of lexical parameter.

However, contrary to Baker’s claims, there is no reason to assume that lexical parameters consistent with the Borer-Chomsky Conjecture are *necessarily* microparametric in their cross-linguistic effects. In fact, as noted in Roberts & Holmberg (2010), «although the idea of reducing parameters to formal features of functional heads has largely been associated with “microparametric” approaches, it is not necessarily restricted to that case» (Roberts & Holmberg 2010, p. 35). Such an association would indeed be in contradiction to what Baker himself argues when he notes that even lexical parameters can have a substantial impact on the grammatical system, thus potentially differentiating even typologically distant languages (cf. Baker 2008a, p. 4). Moreover, if the idea that all parameters do necessarily account for a set of multiple syntactic properties is really taken seriously, since all parameters are assumed to “clusterize” without any exception there is thus neither any epistemological difference between micro- and macro-parameters nor, as pointed out by Kayne’s (2005) analysis, any means of determining it. Curiously enough, an example of such a scenario is provided by Baker himself while pointing out the inherent limits of the microparametric approach to language variation. In this particular case, Baker’s argument considers the difficulty of determining the supposed parametric differences existing between two languages both closely-related and superficially similar:

The shared history of (for example) Fiorentino and Trentino and their on-going interaction guarantees that the two *E*-languages must be similar – that they have roughly the same strings of grammatical words. It does not, however, guarantee that the *I*-languages (internalized grammars; languages seen as a set of rules and principles) are similar. It is possible that two dialects could differ in a macroparameter, but the differences are largely cancelled out on the surface by a series of microparametric choices that the languages make. In that case, the languages would look similar on the surface, and even be mutually intelligible, but have very different parametric structures. [...] I don’t know how serious a problem this possibility presents in practice, but it is worth bearing in mind that there is no guarantee that languages with similar-looking grammaticality patterns always have similar grammars. (Baker 2008a, pp. 6-7)

This excerpt clearly shows the most fundamental aspect of the inconsistency of the distinction between macro- and micro-parameters, namely, that in the above quotation (as well as throughout all Baker (2008a)) these two terms are *de facto* equivalent to the

notions of ‘grammatical parameter’ and ‘lexical parameter’ respectively. If the notions of ‘macroparameter’ and ‘microparameter’ actually referred to the fact that syntactic parameters divide themselves into two classes depending on whether they specifically differentiate either closely-related grammars or typologically distant ones and, most importantly, this difference directly depended on an aspect of their very nature other than their locus of variation, the use of the term ‘macroparameter’ would actually be meaningless in this context. In fact, according to the very definition of microparameter formulated in Kayne (2005), any parameter differentiating two very closely related languages would in principle no longer be a macroparameter, but a microparameter (cf. Kayne 2005, p. 7). Of course, Baker’s argument is completely reasonable and well grounded in stating that the grammatical effects of a parameter might potentially be obscured by the accumulation of numerous others – especially considering the idea of a parameter hierarchy. However, the point here is that the distinction between macro- and micro-parameters would be meaningful only if it referred to an objective, principled explanation justified «on *internal* grounds, on grounds of its relation to a linguistic theory that constitutes an explanatory hypothesis about the form of language as such» (Chomsky 1965, p. 27). Since there is nothing neither objective nor explanatory valid in a purely arbitrary distinction made on such a relative dimension as the possible extent of linguistic variation, this is probably the reason why such an overlapping between the notions of ‘macroparameter’ and ‘microparameter’, on the one hand, and those of ‘grammatical parameter’ and ‘lexical parameter’, on the other hand, has occurred in parametric theory.

### 3.6 – The state of the art of the parameter-setting model for typological variation

In the light of the irreconcilable differences between the macro- and the micro-parametric approaches to parametric variation and their respective shortcomings, Newmeyer’s critical stance on the notion of parameter showed that the revisions of the traditional P&P model as upheld by Kayne and Baker’s accounts were still lacking, especially if taken singularly. However, rather than being sterile and perfunctory, some of the criticisms which had been made in this sense had the merit of precisely pointing out the vulnerabilities of the P&P model as a whole and, as a consequence, of urging the need for a refinement of the concept of parameter which eventually led to the supersedence of the opposition between macroparameters and microparameters.

### *3.6.1 – Against a rule-based model of cross-linguistic variation*

While in his critique of parameters Newmeyer put forth some arguments showing that, at their present stage of development, parameter approaches «have failed to live up to their promise», his proposal for adopting a rule-based model of cross-linguistic variation was generally not well received by generative linguists as it was deemed to represent a retreat to descriptive adequacy, thus denying the commitment to the explanatory adequacy of linguistic theory which has consistently been advocated through the Chomskyan program since the Sixties (Newmeyer 2004, p. 181). In this regard, a notable rejection of Newmeyer's position against the parametric approach to grammar is formulated in Roberts & Holmberg (2005).

In their paper, Roberts & Holmberg respond to each and every point which was raised in Newmeyer (2004, 2005) and which has been presented in section 3.4.2. Concerning the lack of descriptive simplicity of the notion of parameter over that of rule, which was argued by Newmeyer by means of reformulating the Head Directionality parameter as two mutually exclusive language-particular rules, according to Roberts & Holmberg such example shows that «Newmeyer's theory includes rules which are equivalent to parameter settings», thus it is «not formally simpler than standard principles-and-parameters theory» (Roberts & Holmberg 2005, pp. 539-540). Therefore, the point made by Roberts & Holmberg in this regard is that Newmeyer's argument is purely terminological, as the rule-based model advocated by him would not actually add anything to the formal simplicity of the standard P&P approach to typological variation.

The second aspect of Newmeyer's critique of the P&P model which is addressed in Roberts & Holmberg (2005) is the one concerning the binary nature of parameters, which Newmeyer dismissed by referring to the parametrization of the notion of governing category by Manzini & Wexler (1987). Although on the one hand the governing category parameter is surely not binary in its original formulation, the solution adopted by Roberts & Holmberg is that even non-binary parameters can be restated as a hierarchy of multiple binary choices, as shown in (14) (Roberts & Holmberg 2005, p. 540):

- (14)a. Is the binding domain determined by Infl? YES/NO
  - b. If NO, is the binding domain determined by Tense? YES/NO
  - c. If NO, is the binding domain determined by referential Tense? YES/NO
  - d. If NO, is the binding domain determined by root Tense? YES/NO

In these terms, while «the question of binarity is arguably more a matter of formulation than anything else», according to Roberts & Holmberg what actually counts is not so much binarity itself but the idea behind it, that is, the fact that possible cross-linguistic variation is restricted within certain limits which make it systematic and, therefore, always predictable (Roberts & Holmberg 2005, p. 541).

Continuing with their defence of the notion of parameter, the third aspect of parametric theory dealt with in Roberts & Holmberg (2005) concerns the difficulty that would be caused by presupposing too large a number of parameters to be plausible from a language learner's perspective. In addition to pointing out that the number «in the thousands (or, worse, millions)» advocated by Newmeyer would not be needed at all in order to account for the set of all possible grammars, according to the authors what is really important is not the actual number of syntactic parameters, but rather whether these choice-points represent points of underspecification with respect to the initial state of UG (Newmeyer 2004, p. 196). Only if it is answered positively, in fact, does parametric theory conform to the evolutionarily plausible, parameter-free UG assumed by the SMT:

If SMT held fully, which no one expects, UG would be restricted to properties imposed by interface conditions. A primary task of the MP is to clarify the notions that enter into SMT and to determine how closely the ideal can be approached. Any departure from SMT – any postulation of descriptive technology that cannot be given a principled explanation – merits close examination, to see if it is really justified. (Chomsky 2008, p. 135)

Viewing language as essentially «an optimal way to link sound and meaning», any other component attributed to UG would represent an exception to its minimal purpose of being an optimal solution to interface conditions, thus requiring some further explanation rather than being explanatory (*ibidem*). In these terms, assuming with Hauser, Chomsky & Fitch (2002) that the faculty of language in the narrow sense (FLN) – that is, «the only uniquely human component of the faculty of language» – «comprises only the core computational mechanisms of recursion as they appear in narrow syntax and the mappings to the interfaces» (Hauser *et al.* 2002, pp. 1569, 1573), such a restricted view of FLN excludes the possibility of having an overspecified UG providing a full array of discrete parametric options, while still potentially allowing for parametric variation in the lexicon.

Strictly connected to the minimalist conception of UG is the existence of parameter hierarchies of the kind proposed by Baker (2001), which Newmeyer challenged mainly on

empirical grounds. While admitting the difficulty of Baker's endeavour, Roberts & Holmberg argue that the alternative proposed by Newmeyer would not fare any better, as «language-specific rules make absolutely no prediction about complex patterns of relatedness of the type Baker's PH tries to make» at both the empirical and the theoretical level (Roberts & Holmberg 2005, p. 542). This is due to the fact that, assuming language-specific rules to have different properties from syntactic parameters and hence not to simply amount to mere notational variants of these latter, we would have a completely unconstrained, continuous spectrum of linguistic variation rather than a discrete one, the reason being that such rules would nonetheless still require some kind of format on their formulation:

As things stand, [language-specific rules] predict that languages may vary without assignable limits. They may lack the empirical difficulties that beset something like Baker's PH, because, lacking any constraint on their formulation of any kind, they are able to predict everything and therefore nothing. (*ibidem*)

The last part of Roberts & Holmberg's reply to Newmeyer's criticisms discusses the existence of abstract implicational relations between individual parameter settings. This point is arguably the most relevant to the present work, as if on the one hand such deductive structure should allow the prediction of parametric clusterings of syntactic properties in the spirit of Chomsky (1981a), on the other hand, as notably pointed out by Newmeyer for Rizzi's (1982) Null Subject Parameter, such predictions have not been borne out by empirical research. Leaving aside for later discussion (3.6.2) the strong empirical strand in their treatment for this parameter, the authors argue that Gilligan's (1987) extensive linguistic data still suggested the existence of the following four implicational statements (Newmeyer 2005, p. 90):

- (15)a. NULL TS → NULL NTS
- b. SI       → NULL NTS
- c. SI       → THAT-T
- d. THAT-T → NULL NTS

While the correlations listed in (15) were described by Newmeyer as «not very heartening [...] for any which sees in null subject phenomena a rich clustering of properties», according to the authors such results do not diminish the value of Rizzi's proposal (*ibidem*). In fact, even if the parametric clustering of syntactic properties originally

proposed by Rizzi (1982) does not hold, his intuitions regarding the existence of a link between some two seemingly independent properties as the licensing of null subjects in finite clauses and the lack of *that*-trace effects in embedded clauses «were clearly on the right track» (Roberts & Holmberg 2005, p. 544):

[...] consider what has actually been shown: an original very strong correlation was postulated on the basis partly of theoretical considerations and partly the close comparison of a small number of closely related languages. When a very large number of genetically and typologically highly diverse languages were compared for the 'same' properties, with no control as to the other typological features of these languages, the original correlations were shown not to hold in their original form, although four implicational statements could still be gleaned and five unsuspected language types observed. To us, this does not seem like a bad or shocking result for parametric theory, but rather a fairly promising result from the admixture of a very large amount of essentially random data into an originally carefully controlled database. (*ibidem*)

Finally, according to Roberts & Holmberg, «the postulation of language-specific rules in place of parameters would have revealed precisely nothing either in the controlled database studied by Rizzi or in the random sample chosen by Gilligan»; hence, the superiority of the P&P model for linguistic variation over a purely descriptive rule-based one, despite the former being still in need of further theoretical and practical development (*ibidem*).

In conclusion, Roberts & Holmberg (2005) represents a clear stance against the adoption of a rule-based model of typological variation in place of the standard principles-and-parameters theory. Crucially, however, in this context the superiority of the notion of parameter in accounting for the possible range of syntactic variation is not absolute, but only relative to explanatory adequacy – that is, an ideal minimal standard which whichever set of observationally adequate rules would never manage to reach due to their unconstrained descriptive power. In this sense, therefore, if on the one hand a highly abstract modular system of principles and parameters will always in principle perform better than a set of language-particular rules in explaining how language faculty works, the disaggregation of all alleged (macro)parameters into arrays of (micro)parameters equivalent to merely descriptive statements, together with the evolutionary implausibility of the traditional, over-specified conception of UG, still undermined the P&P model at this theoretical stage.



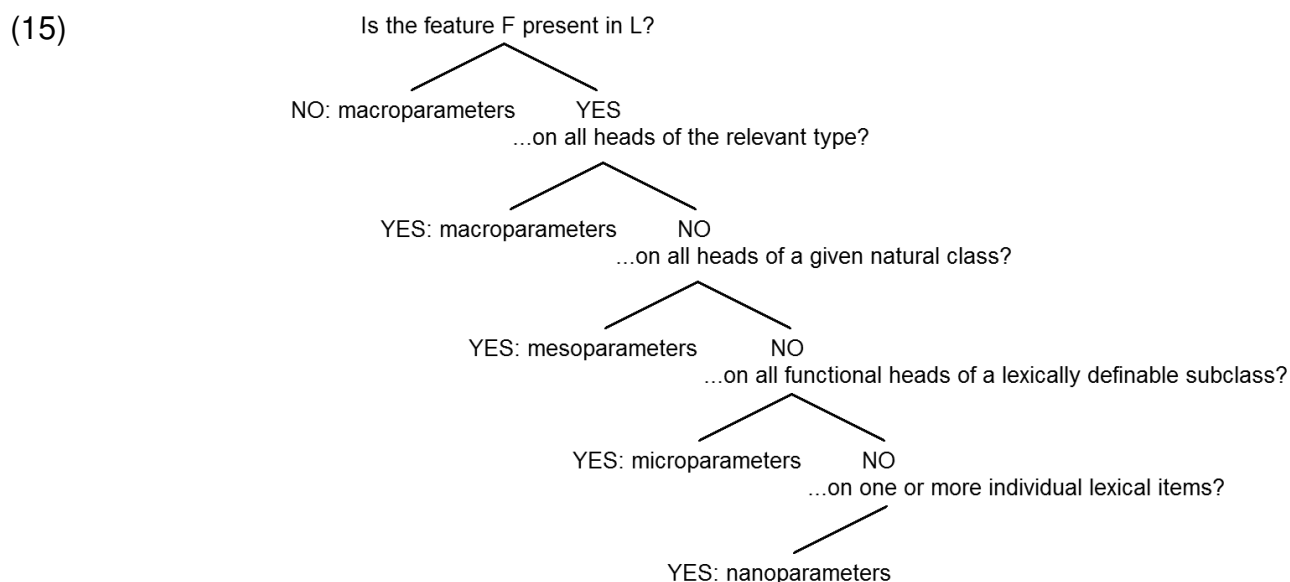
### 3.6.2 – *Roberts and Holmberg's hierarchical model of parametric variation*

By balancing «the optimistic, difficult, abstract path» represented by the parameter-based theory with a purely rule-based one, it is clear that only the former can have some sort of predictive power and, accordingly, can represent a way to overcome the tension between descriptive and explanatory adequacy raised by Chomsky (1981a) (Roberts & Holmberg 2005, pp. 551-552). However, the superiority of the concept of parameter over that of rule in this respect has mainly been nominal, as in practice a solution to such tension has never been provided in full; neither by the macroparametric model, nor by the microparametric one. On the one hand, the fact that the macroparametric approach still referred to the obsolete conception of parameterized UG made it unable to suit into the invariant, rudimentary biological view of language faculty of contemporary Generative Grammar. On the other hand, the progressive exaggeration of the descriptive power of modern microparameters has been at the detriment of their explanatory power, thus making them roughly equal to taxonomic devices devoid of any actual foundation with respect of a principled theory of language acquisition. For all these reasons, then, although at the end of the 2000s the concept of parameter was still worth defending, the already existing model of parametric variation needed to be rethought from the ground up.

Generally speaking, while what Roberts & Holmberg did in their reply to Newmeyer was to emphasize the inherent strengths of the parametric enterprise and their perceived lack of any possible alternative to such line of research, only five years later would a new, viable parametric model be actually proposed. In these terms, Roberts & Holmberg (2010) (and its further developments such as Roberts (2012) and Biberauer & Roberts (2015)) represents both the best solution yet found to the problems raised by Newmeyer (2004, 2005). This model, which is referred to as «hierarchical» by the authors themselves, combines the underspecification view of language faculty put forth in the Borer-Chomsky conjecture with the more traditional idea that parameter settings are arranged in a hierarchical fashion, thus representing a sort of intermediate approach to the question of micro vs macroparametric variation (Roberts & Holmberg 2010, p. 2). However, what really sets it apart from previous parametric approaches is that due attention is paid not only to the role of UG and primary linguistic data in language acquisition and possible typological variation, but also to that of third factor considerations in the sense of Chomsky (2005) – that is, the set of general considerations of computation and cognition which are shared by (but not strictly specific to) language faculty (cf. Chomsky 2005, p. 6):

The third factor falls into several subtypes: (a) principles of data analysis that might be used in language acquisition and other domains; (b) principles of structural architecture and developmental constraints that enter into canalization, organic form, and action over a wide range, including principles of efficient computation, which would be expected to be of particular significance for computational systems such as language. (Chomsky 2005, p. 6)

According to Roberts & Holmberg (2010), all of this can be achieved by «retaining a formally “microparametric” view of macroparameters, i.e. seeing macroparameters as aggregates of microparametric settings» but at the same time upholding the idea that «these aggregate settings are favoured by markedness considerations», which in turn reflect the acquirer’s conservative tendency to generalize any feature (of a given value) F specified on a functional head of a given type to all comparable heads in a target language L (Roberts & Holmberg 2010, p. 39). In this sense, interdependencies between microparametric settings are hierarchical in such a way that a positive setting for a low-level, more marked parametric option necessarily depends on a specific setting for all its higher-level, less marked parametric options, but not vice versa, as shown in (15) (cf. Biberauer & Roberts 2015, p. 302):



In this instance, special reference must be made to the fact that the terms *macroparameter* and *microparameter* (together with *mesoparameter* and *nanoparameter*) adopted in the hierarchical template shown above do not directly refer to either any possible approach to cross-linguistic comparison nor Baker’s (now dismissed) distinction between grammatical parameters and lexical parameters. Being based on a view of parametric variation exclusively restricted to the lexicon, Roberts & Holmberg’s (2010)

hierarchical model for typological variation not only avoids the fallacy of arbitrarily assuming some parameters to be inherently be more prone than others to account for large-scale (or small-scale) typological differences, but also accounts for how parametric settings can potentially have the same descriptive power of language specific-rules and still be, to a certain degree, explanatorily significant. This particular aspect is precisely dealt with in the following scalar typology of parameter settings:

(16) For a given value  $v_i$  of a parametrically variant feature F:

- a. Macroparameters: all heads of the relevant type, for example, all probes, all phase heads, and so forth, share  $v_i$ ;
- b. Mesoparameters: all heads of a given natural class, for example, [+V] or a core functional category, share  $v_i$ ;
- c. Microparameters: a small, lexically definable subclass of functional heads, for example, modal auxiliaries, subject clitics, etc. shows  $v_i$ ;
- d. Nanoparameters: one or more individual lexical items is/are specified for  $v_i$ ;

(Biberauer & Roberts 2015, p. 302)

According to (16), every hierarchical schemata of the same type as (15) refers to an individual parameter, here meant as a specific formal feature F which is assigned by the learner to a more or less large subset of functional heads. In these terms, the labels *macroparameter*, *mesoparameter*, *microparameter* and *nanoparameter* do not refer to distinct parameters, but rather to the possible degree of spreading of a given feature in the grammatical system. Crucially, the subset of functional heads which is specified for such feature is retrieved by the learner by relating, on a case-by-case basis, relevant pieces of primary linguistic data (the second factor) with his own innate endowment for language (the first factor) by means of a general cognitive inferential process – originally formulated as *Generalization of the input* in Roberts (2007) – expressing the general requirement «to “make maximal use of minimal means”» (Biberauer & Roberts 2015, p. 300):

(17) Input Generalisation (adapted from Roberts 2007: 275):

If a functional head F sets parameter  $P_j$  to value  $v_i$  then there is a preference for all functional heads to set  $P_j$  to value  $v_i$ .

(*ibidem*)

In addition to corresponding to a third-factor consideration, this heuristically overgeneralizing procedure can also be justified «due to ignorance of categorial distinctions» of the acquirer, which «gradually erodes through the learning process [...] as

finer and finer distinctions are made as a consequence of the interaction of all three factors» (Biberauer & Roberts 2015, p. 301). In these terms, if on the one hand maximum feature generalization corresponds to a minimally marked system, on the other hand «each more deeply embedded option is more marked than all less deeply embedded ones» as it requires the postulation, on the part of the learner, of some further linguistic hypothesis of added descriptive complexity to approximate the target grammar (Roberts & Holmberg 2010, p. 46):

So here we see a parameter schema given as a network of options, each more embedded option representing a more specific, and therefore a more marked, option. [...] again the conservatism of the learner is such that it prefers the path to be as short as possible, and so deeply embedded options are relatively marked owing to the fact that they have longer descriptions. Following GGL [Gianollo, Guardiano & Longobardi 2008], we assume that the schema and the overall pool of possible features are given by UG; the network is created through epigenesis in acquisition, and markedness follows, on one standard construal, from increasing specificity [...]. (Roberts & Holmberg 2010, p. 47)

As shown in (15), according to Roberts & Holmberg the maximally unmarked (macroparametric) option for the grammatical system is selected when «'no F has this value'», that is, in case no functional head in the grammatical system is specified for the given formal feature (*ibidem*). Then, the next least marked (macroparametric) option is selected when all functional heads of the relevant type are specified for F with its value set to  $v_i$ , while non-harmonic and thus more marked systems relate the possession of F to heads of a particular subclass requiring either to be specified for additional categorial features (mesoparameters) or to be further defined also in terms of lexical categories (microparameters) (cf. *ibidem*). Finally, at the bottom of the hierarchy there are those configurations which the parameter schemata in (15) refers to as *nanoparameters* – namely, «synchronically systematic options that affect only a very small number of lexical items (possibly only one)» (Biberauer & Roberts 2015, p. 302). Although the term nanoparameter has been introduced in Boeckx (2008) in order to denote indifferently all syntactic parameters, which in his non-hierarchical system of linguistic variation are regarded as «parametric options isolated from one another and localized to specific head», in the context of Roberts & Holmberg's hierarchical model nanoparameters actually do not correspond to parameters in the true sense of the word (Boeckx 2008, p. 216). In fact, while being part of the same hierarchy of macro-, meso- and microparameters, nanoparameters are acquired independently from higher-level parametric settings as they

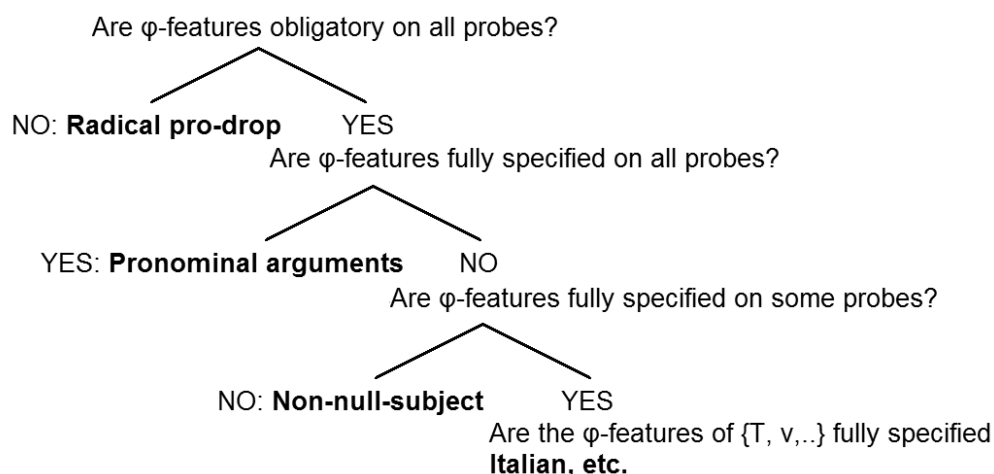
represent, with respect to the specific grammatical system in which they are hosted, an unsystematic set of lexical idiosyncrasies which have to be acquired on a case-by-case basis. As stated in Biberauer & Roberts (2015):

Many, perhaps all, nanoparametric options fall outside the core system defined by the parametric hierarchies under discussion here. To the extent that nanoparametric options involve high-frequency elements, they appear to be acquired as independent lexical items, independently of the more general properties of the system to which they belong; [...] In our terms, forms of this type would therefore not be acquired as a result of progressing down a given hierarchy, although their connection to specific hierarchies – in the sense that they appear to represent isolated instantiations within a given system of a pattern that can be seen to hold more systematically in other systems – is clear. (Biberauer & Roberts 2015, p. 303)

Returning now to the crucial tension between the epistemological role of syntactic parameters in a theory of language acquisition and the increase in their descriptive power associated with the formulation of the BCC, what emerges from the present outline is that, according to Roberts & Holmberg's model, parametric status is attributed exclusively to those descriptive statements whereby a given feature *F* is, by means of the general inferential procedure in (17), *generalized and generalizable* to a systematic set of functional heads (however big or small the set is), with all other cases of lexical feature specification falling outside what Chomsky & Lasnik (1977) refer to as «a theory of core grammar with highly restricted options, limited expressive power, and a few parameters» (Chomsky & Lasnik 1977, p. 430). Therefore, accordingly, as this format of parametric variation now excludes any element which cannot be fitted into all the possible deductive interactions left unspecified by UG and third factor considerations, all those unsystematic, isolated properties eluding such heuristic path have not to be regarded as parameters, but rather as the expression of maximally descriptive and minimally explanatory language-specific rules which are independent from the parametric space.

As an example of their proposed parameter model, all the above considerations are applied by Roberts & Holmberg to the formulation of the Null Subject Parameter. Although the basic form of this parameter was first proposed in Roberts & Holmberg (2010), the formulation proposed here is taken from Roberts (2012) for its greater concision (cf. Roberts 2012, p. 324):

(18)



As pointed out in (18), the hierarchy associated to the Null Subject Parameter yields four basic linguistic types. Radical pro-drop languages are obtained by selecting the unmarked option; these correspond to languages like Japanese, whose most salient property with respect to null argument licensing is to «lack agreement-marking altogether and yet permit any pronoun to be dropped under appropriate discourse conditions» (Roberts 2012, p. 324).

The next least marked option gives rise to languages which, like Navajo, are grouped together under the label “pronominal argument systems” – namely, languages which «allow all pronominal arguments to drop, but differ from the East Asian type in showing fully specified subject- and object-agreement, as well as possessor agreement in nominals, and free word order» (*ibidem*). Assuming with Roberts (2010) that «null arguments [...] arise through pronoun deletion, which can take place under the generalized recoverability condition that the formal features of the goal be (properly) included in the features of the probe», in grammatical systems belonging to this type it is precisely the generalized presence of unvalued  $\phi$ -features on all probes which allows pro-drop to potentially affect all of the verb’s pronominal arguments (*ibidem*).

Going downwards in the hierarchy to non-homogeneous systems, the negative value of the binary setting distinguishing the distribution of unvalued  $\phi$ -features on either no probes or some syntactically defined subsets of functional heads yields, respectively, non-null-subject languages like English and languages which, like Italian, allow only the licensing of null pronominal subjects (cf. Roberts 2012, p. 325). Although the schema in (18) is not directly concerned with a typology of null-subject languages, in Roberts & Holmberg (2010) it is argued that consistent null-subject languages like Italian do not exhaust all parametric possibilities. In fact, as it will be more thoroughly discussed in

Chapter IV, at this point in the hierarchy «the null-subject parameter starts to ‘break up’ into microparameters as individual probes are evaluated in relation to it» (Roberts & Holmberg 2010, p. 49).

In conclusion, compared to the conceptions of parametric variation represented and argued for by Baker (2001, 2008) and Kayne (2000, 2005) respectively, the model proposed by Roberts & Holmberg (2010) represents the state of the art of the P&P theory. Its advantage over its two ideal predecessor consists, more than anything else, in the fact that attention is not only focused on the role of UG and primary linguistic data but also on the role of third-factor considerations, which crucially take away some of the epistemological burden from the former and systematize the triggering effects provided by the latter. In these terms, therefore, parametric variation can be taken to be nothing more but «an emergent property of the interaction between the learner, the primary linguistic data, and UG» (Roberts 2012, p. 320).

### 3.7 – Conclusion

After focusing on the technical aspects of the latest development of the P&P model, I conclude my discussion by taking a broader perspective on what may be argued to be the main factors characterizing the debate about the concept of parameter which took place during the first decade of the 21<sup>st</sup> century. While being sequentially arrangeable, these factors respectively represent theory-internal considerations, theory-external objections and what could be ideally be regarded as their joint outcome. The first aspect is directly linked to the fact that, although cross-linguistic variation had already been assumed to be restricted to formal features of lexical items from almost the beginning of the P&P framework, the microparametric conception entered a major debate against the more traditional, macroparametric one only twenty years after its original formulation in the Borer-Chomsky conjecture. In these terms, what emerges is that the idea of a direct parametrization of UG was still not completely abandoned with the advent of Minimalism. This is possibly due to the fact that the rich deductive structure traditionally attributed to macroparameters, that is, that very aspect which allowed linguistic theory to attain predictive power for both language variation and acquisition in the spirit of Chomsky (1981a), still had – and still has nowadays – a great appeal in Generative Grammar. Having said this, what is worth noting is that while on the one hand the necessity of decomposing the traditional parameters of the GB Theory into lexical microparameters

was initially determined by empirical, typological reasons, it is nonetheless clear that the notion of macroparameter as embedded in the traditional P&P model could not be reconciled with minimalist demands. Therefore, the abandonment of grammatical parameters in the sense of Baker (2008a) was not a consequence of growing empirical evidence against macroparametric clusters, but rather an effect of the theoretical development of Generative Grammar itself. Second, it is interesting to note that Roberts & Holmberg (2010), in their attempt to reformulate a notion of parameter which is consistent with minimalist assumptions, reclaimed for their hierarchical model exactly those theoretical aspects which had been pointed out by Newmeyer (2004, 2005) as actually not being consistently upheld within the current parametric approaches – most notably, the lack of both a strictly binary format in some traditional syntactic parameters and a general intent of pursuing the acknowledged, although not unproblematic, aim to «working out the relevant implicational relations among parameters» (Newmeyer 2004, p. 196). In this sense, while not proposing a viable alternative from the point of view of a linguistic approach refusing a continuous, unrestricted view of language variation, it can be argued that the claims put forth by Newmeyer contributed, although indirectly, to the exclusion of some inconsistencies plaguing previous parametric approaches from Roberts & Holmberg's own model of linguistic variation. Finally, by looking at Roberts & Holmberg's (2010) and Biberauer & Roberts's (2015) attempt to reformulate a notion of parameter which could be consistent with minimalist assumptions, the fact itself that parametric variation is now regarded as an emergent property of the interaction of UG, primary linguistic data and third factor considerations crucially implies that «the ontological status of the parameters, and of the hierarchies, is very different from either the 1980s or the 1990s conception» (Roberts 2012, p. 322). In particular, parameter hierarchies are not encoded as either autonomous syntactic entities nor overspecified aspects of UG and thus exist only as derived concepts, while those entities which have been referred to as syntactic parameters are finally regarded as feature specifications on lexical items. In these terms, the question of whether or not «as Holmberg and Robert themselves acknowledged [...], "the notion of parameter is almost empty; it really doesn't have much content"» is exactly the one which will be taken into further analysis in the following chapters (Boeckx 2010, p. 13).



## Chapter IV

### The main parameters of the GB Theory in current generative theory

#### 4.1 – Some preliminary considerations on Rizzi's (2014) list with respect to modern generative theory

Considering the parameters included in the list proposed by Rizzi (2014) and retrospectively classified in Chapter II according to the subdomains of today's Generative Grammar, an analysis of their current status requires, first of all, such list to be amended by taking out those parameters which, due to the developments within generative theory itself, have been shown either to refer to obsolete theoretical concepts or to be reducible to other, more basic theories.

(1) The main syntactic parameters of the GB phase in current generative theory (revised):

- (I) Locality parameters:
  - i. ~~bounding nodes~~
  - ii. ~~long-distance anaphors~~
- (II) Case-assignment parameters:
  - i. ~~NOM assigned by means of either agreement or government~~
- (III) Merge parameters:
  - i. ~~"believe" and S' deletion~~
- (IV) Linearization parameters:
  - i. ~~X-bar vs W\* languages~~
  - ii. *head-complement parameter*: Kayne (1994), Richards (2004, 2008), Biberauer, Holmberg & Roberts (2009, 2014), Biberauer & Roberts (2015);
- (V) Spellout parameters:
  - i. *null subject*: Roberts (2010), Holmberg (2010a);
  - ii. *V movement to I*: Biberauer & Roberts (2010, 2012), Holmberg & Roberts (2013);
  - iii. *V movement to C*: Roberts (2004);
  - iv. ~~P-stranding~~: reduced to (*abstract*) *incorporation*;
  - v. (*abstract*) *incorporation*: Baker (1996), Biberauer *et al.* (2014);
  - vi. *Overt vs covert wh-movement*: Richards N. (2010);

As shown in (1), the parameters which are dealt with in the present section are the ones which were assumed to account for, respectively, the patterns of cross-linguistic

variation in the choice of bounding nodes, the possibility of long-distance anaphora, the assignment of nominative Case, and, finally, the typological distinction between configurational and non-configurational languages. In this respect, special reference should here be made to the Merge parametric class as it will not be referred to in the next section. As far as the lexical parameter of S'-deletion is concerned, assuming with Chomsky (1986b) that «the exceptional Case-marking property of *believe*-type verbs [...] in English must be specifically learned», this proposal as an instance of parametric specification can be disregarded in the present discussion since it represents, on the other hand, a property associated to a very specific set of lexical items which would not be ascribed to any parametric option in the true sense of the word (Chomsky 1986b, p. 190). In fact, this statement is true even assuming, in accordance to Biberauer & Roberts (2015), that this property is dependent on a feature localized to a specific head, as in this case it would still correspond to a nanoparameter – that is, a non-parametric option.

Finally, due to its particularly debated status within the development of Generative Grammar and its relevance for the purpose of the present study, the comprehensive history of the head-complement parameter is reviewed separately in Chapter V.

#### 4.1.1 – *The parametrization of bounding nodes, the Barriers framework and the Phase Impenetrability Condition*

As previously mentioned in Chapter II in connection with Baker (1988) and Pollock (1989), within the framework of Chomsky (1986a) the notion of *bounding node* referred to by Subjacency was replaced with the notion of *barrier* which, unlike its theoretical predecessor, could not be specified in terms of parametric variation. In fact, not only were barriers «relative [...] to the potentially governed element» (Baker 1988, p. 56), thus allowing any XP, regardless of its category, to act as a barrier if intervening between any pair of links in a movement chain, but also the notion of barrierhood itself was not held as an absolute. This latter development was in fact due to the notion of barrierhood being indirectly defined in terms of L-marking, a specific condition of sisterhood allowing a potential barrier to be exempted from barrierhood, as originally proposed in (2-3):

(2)  $\gamma$  is a barrier for  $\beta$  iff (a) or (b):

- a.  $\gamma$  immediately dominates  $\delta$ ,  $\delta$  a BC [BLOCKING CATEGORY] for  $\beta$ ;
- b.  $\gamma$  is a BC for  $\beta$ ,  $\gamma \neq \text{IP}$ .

- (3)  $\gamma$  is a BC [BLOCKING CATEGORY] for  $\beta$  iff  $\gamma$  is not L-marked and  $\gamma$  dominates  $\beta$ .  
(Chomsky 1986a, p. 14)

As Chomsky (1986a) replaced bounding nodes with barriers, within Minimalism the notion of barrier itself has been in turn replaced by the notion of *phase*. According to the so-called Phase Theory first outlined in Chomsky (2000, 2001b), phases are the minimal units of syntactic derivation which are cyclically transferred from narrow syntax to the interfaces. More specifically, each phase corresponds to a chunk of *lexical array* (LA), the latter being meant as a one-time selection of lexical items which, in order to reduce the computational burden for the speaker during syntactic derivation, does not require further access to the lexicon until the derivation itself is completed (cf. Chomsky 2001b, p. 11):

Suppose we select [a lexical array] LA [...]; the computation need no longer access the lexicon. Suppose further that at each stage of the derivation a subset  $LA_i$  is extracted, placed in active memory (the “work space”), and submitted to the procedure L [computation/narrow syntax]. When  $LA_i$  is exhausted, the computation may proceed if possible; or it may return to LA and extract  $LA_j$ , proceeding as before. (Chomsky 2000, p. 106)

In Chomsky (2000, 2001b) the phases are vP and CP, the reasons from this choice mainly deriving from their being «relatively independent in terms of interface properties» (*ibidem*). On the semantic side, if the propositional nature of these maximal projections represents the realization of the requirement that each derivation chunk must be a semantically complete constituent, from a phonological perspective «they have a degree of phonetic independence» which distinguished them from other constructions (Chomsky 2001b, p. 12). However, what is crucial here with respect to the locality conditions on movement once accounted for by Subadjacency is that phases are also characterized by the property of being inaccessible to further syntactic operations *in their internal domain* – that is, in the syntactic configuration corresponding to the complement of the phase head. This generalization corresponds to the Phase-Impenetrability Condition (PIC), a condition on derivation emerging from the need to further reduce the computational burden by allowing the phonological component to «“forget” earlier stages of derivation» (Chomsky 2001b, p. 13). Although the PIC was first proposed in Chomsky (2000), the formulation adopted here is taken from Boeckx & Kleanthes (2004) because of its greater clarity:

(4) Phase Impenetrability Condition:

Once a phase has been completed and sent to the interfaces, the internal domain of a phase [i.e. the complement of the phase head] is not accessible to operations at/above the next higher phase. Only the edge of the phase [the head plus any number of specifiers] remains accessible at the next higher phase.

(Boeckx & Kleanthes 2004, p. 1)

In these terms, leaving aside the still relevant conceptual shift carried out in the *Barriers* framework, in Minimalism the role of Subjacency in accounting for locality constraints in long distance syntactic movement has finally been taken over by an independent set of minimal design specifications which, in turn, are currently assumed to be related to both legibility conditions imposed by external systems and general computational considerations.

#### *4.1.2 – A minimalist account of long-distance anaphora: the logophoric use of reflexives*

In the preceding section, the parametrization of bounding nodes proposed by Rizzi (1978) has been shown to have been essentially replaced by the Phase-Impenetrability Condition of Chomsky (2000, 2001b). Although being mainly due to the general development of Generative Grammar from a minimalist perspective, this theoretical reduction represented a theoretical improvement over Rizzi's (1978) original idea also from another point of view as the latter implied, contrary to the later standardly assumed Borer-Chomsky Conjecture, the possibility that principles could be parametrized. Turning now to the governing category parameter put forth in Manzini & Wexler (1987), it can be argued that the dismissal of this specific parameter followed from this very same reason. In fact, if on the one hand the range of cross-linguistic variation in anaphoric dependencies presented in Manzini & Wexler (1987) was effectively analyzed according to the modern idea that «parameters are associated with lexical items in a grammar rather than with grammars as a whole», on the other hand this same proposal assumed the possibility that a subdomain of UG as Binding Theory could be subject to parametric variation (Manzini & Wexler 1987, p. 415). This was an implication of Manzini & Wexler's idea that the syntactic domain in which anaphoric elements are free and bound is not the same for all of them,

but rather is specified from case to case according to the possible domains defined by (a) through (e) in the following definition of governing category:

(5)  $\gamma$  is a governing category for  $\alpha$  iff

$\gamma$  is the minimal category that contains  $\alpha$  and a governor for  $\alpha$  and

a. can have a subject or, for  $\alpha$  anaphoric, has a subject  $\beta$ ,  $\beta \neq \alpha$ ; or

b. has an Infl; or

c. has a Tense; or

d. has a “referential” Tense; or

e. has a “root” Tense;

if, for  $\alpha$  anaphoric, the subject  $\beta'$ ,  $\beta' \neq \alpha$ , of  $\gamma$ , and of every category dominating  $\alpha$  and not  $\gamma$ , is accessible to  $\alpha$ .

(Manzini & Wexler 1987, pp. 422-423)

Similarly to what happened to the locality phenomena once accounted for by the Subjacency parameter, nowadays the range of cross-linguistic variation in long-distance anaphoric binding is not understood in terms of an overspecified UG but is rather regarded as a lexically-dependent property. In this respect, an example of how long distance anaphora is dealt with in current minimalist theories is provided by Sportiche (2013). Building on Pollard & Sag’s (1992) and Reinhart & Reuland’s (1993) extensive work on exempt anaphors – that is, a specific type of reflexives which, under certain syntactic and semantic conditions, «can have remote antecedents and thus need not be subject to Condition A» – in this paper Sportiche argues that, in languages such as English and French, long distance (and non c-command) binding is a property shared by those reflexives whose antecedent is introduced as «a perspective holder, e.g. someone whose words, thoughts, point of view are being reported or perspective adopted» (Sportiche 2013, p. 201). Accordingly, one possibility is therefore that the possibility of long-distance anaphora, here conceived as an exemption from Principle A (or Condition A), is connected, on the one hand, with the logophoricity of some specific reflexives (in the sense of requiring to be anteceded by a perspective holder) and, on the other hand, with a set of specific syntactic conditions to be met by the reflexive with respect to its antecedent – all this without any need to resort a the parametrization of the notion of governing category.

#### 4.1.3 – *The abandonment of the notions of typological non-configurationality and government in Generative Grammar*

Concluding the section about those of the parameters of the GB Theory which have not stood the test of time, the last parameters reviewed in this respect are the parameterization of the Projection Principle and the parameter of nominative case assignment. Differently from what has been written above for the other “classical” parameters which did not find a place within Minimalist syntax, these two parameters owe their abandonment mainly to the fact that (at least) one of the core concepts on the basis of which they had been respectively formulated has later been deemed theoretically inadequate. As far as the parametric proposal by Hale (1983) is concerned, the first and foremost factor which determined its abandonment has been the overcoming of the original distinction between configurational and non-configurational-languages as a distinction between two basic natural classes. This need for change was remarked also by Hale himself, who in his paper *On nonconfigurational structures* (1986) recognized not only that, contrary to what is commonly held since the publication of Chomsky (1981a), «nonconfigurationality is not a global property of languages» but rather «a property of constructions», but also that the Configurationality Parameter proposed in Hale (1983) «made languages seem more different than, ought to be possible» in that it had them differing in relation to the Projection Principle, a *fundamental* aspect of the grammars of all languages» (Hale 1986, p. 352):

[...] the term «configurationality» is not a particularly appropriate one, since the notion «configuration», in the sense of a hierarchical organization of constituents is essential to all languages (certainly at the LS representation, but arguably also at PS). (*ibidem*)

As the notion of non-configurationality became a general descriptive label indicating those languages whose underlying hierarchical structures are not readily available at surface level due to the interference of other grammatical factors, the idea of a parameter constraining the mapping between each level of representation had no more reason to be taken into account. Accordingly, although in later times both D-Structure and S-Structure – and hence the Projection Principle itself – would also be eliminated from Generative Grammar altogether, the fall of Hale’s (1983) Configurationality Parameter can be argued to have been primarily a consequence of the obsolescence of «a typological distinction

which too drastically opposed languages with nonconfigurational structures to languages which, for the most part, lack them» (Hale 1989, p. 300).

Regarding the dispensing of the parameter of nominative case assignment, the issue at stake here is certainly the notion of government, here re-proposed according to Chomsky's (1981a) formulation (which in turn was conceived as a redefinition of Aoun & Sportiche's (1981) earlier formulation):

(6) [ $\beta \dots \gamma \dots \alpha \dots \gamma \dots$ ], where

(i)  $\alpha = X^0$

(ii) where  $\phi$  is a maximal projection, if  $\phi$  dominates  $\gamma$  then  $\phi$  dominates  $\alpha$

(iii)  $\alpha$  c-commands  $\gamma$

(Chomsky 1993 [1981a], p. 165)

According to (6), government consists basically in a local version of c-command (cf. Chomsky 1995a, p. 35). As such, however, its theoretical status was highly dubious as not only did this concept have to be taken as an axiom but, equally undesirably for a research program questioning all those theoretical devices which could not to meet language faculty's minimal design specification, it also coexisted with the simpler/more general notion of c-command. On top of that, government itself subsumed a whole host of more specific syntactic relations, each of which was crucially involved in a distinct subsystem of grammar; for example, *head government* (the proper government of GB Theory) was a derivative notion which accounted for both the licensing of empty categories and Case assignment (cf. Chomsky 1995a, p. 82). For all these reasons, the Minimalist Program has abandoned government theory altogether in favour of that minimal set of «crucial properties and relations» which could be stated «in the simple and elementary terms of X-bar theory», the latter being held by Chomsky (1995a) as the true theoretical primitive in a theory of UG (Chomsky 1995a, p. 172):

The computational system takes representations of a given form and modifies them. Accordingly, UG must provide means to present an array of items from the lexicon in a form accessible to the computational system. We may take this form to be some version of X-bar theory. The concepts of X-bar theory are therefore fundamental. [...] The version of a minimalist program explored here requires that we keep to relations of these kinds, dispensing with such notions as government by a head (head government). (Chomsky 1995a, pp. 172-173)

In these terms, head government has been substituted with the more fundamental *head-complement relation* (or *sisterhood*), that is, «the core local relation», with the other local relation expressible in terms of X-bar theory being the one between a head and its specifier (Chomsky 1995a, p. 173). More crucially for the present purposes, in Chomsky (1995a) all instances of Case assignment were reduced to the Spec-head relation (*agreement* in Koopman & Sportiche (1991)), hence providing a single elegant, unifying theory:

Take Case theory. It is standardly assumed that the Spec-head relation enters into structural Case for the subject position, while the object position is assigned Case under government by V, including constructions in which the object Case-marked by a verb is not its complement (exceptional Case marking). The narrower approach we are considering requires that all these modes of structural Case assignment be recast in unified X-bar-theoretic terms, presumably under the Spec-head relation. (*ibidem*)

Based on such conditions, it is clear that Koopman & Sportiche's (1991) dualistic account of nominative Case assignment could not be reconciled to such an minimalist account. In fact, while on the one hand the notion of government plays no role in Minimalism, thus excluding one of their proposed parametric possibilities, on the other hand the unification of Case assignment under agreement would still eliminate the possibility of specifying a given functional head as either a governed or an agreement Case assigner.

## 4.2 – The main GB parameters in Minimalism

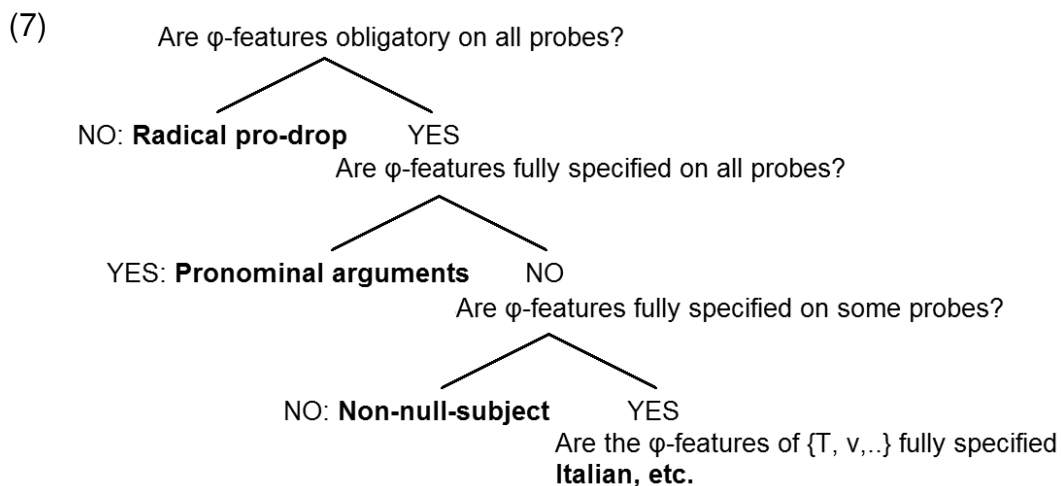
Referring to list (1), this section is devoted to analyzing those GB parameters whose epistemological status is still being upheld in Minimalism. In this respect, for each of them one possible formulation has been selected on grounds of both theoretical interest and, when possible, compatibility with the parametric model outlined in Roberts & Holmberg (2010). As will become clear, what is more relevant to the aim of this thesis is not the specific details of each of these formulations (which, in any case, simply represent some steps in the understanding of some instances of cross-linguistic variation and certainly not the final words on them), but rather whether the general ideas behind them ultimately conform to the general classification which has been proposed in Chapter I.



#### 4.2.1 – The null subject parameter

Since its first and “classical” formulation in Rizzi (1982), the null subject parameter has always had a special place in the P&P model because of both its formulation, which even in its earliest implementation was already complying with the parametric format which would be stated by Borer (1983), and in its deductive character, as it aimed at accounting for not only for a single syntactic property but for a cluster of them. Considering its importance in this respect, this parameter has also been receiving particular attention from both those questioning the plausibility of a parameter-setting approach and those favourable to such a line of research. Notably, the null subject parameter is the central one in Roberts & Holmberg’s (2010) proposal for a model for typological variation compatible with minimalist assumptions, and it is precisely for this reason that the version of the null subject parameter referred to here is the one proposed by Roberts (2010) and Holmberg (2010a).

Although in their respective papers Roberts and Holmberg reach different conclusions about the syntactic features involved in some specific null subject phenomena, these authors often and openly refer to each other’s works as their proposals both rely on the same basic intuition and complement each other in some major respects, with Holmberg’s (2010a) account being seen here as a development, in a typological sense, of Roberts’s (2010) perspective on the nature of “pro-drop”. As far as their shared views are concerned, both the papers referred to here deal with those languages which correspond to the bottom-right branch of the null subject parametric schema in (7) (Roberts 2012, p. 324):



As already mentioned in Chapter III (section 3.6.2), this “mesoparametric” group is not the most marked in an absolute sense as it still allows for varying internal differentiation in terms of further parametric variation. According to Holmberg (2010a), at this point in the hierarchy two varieties of null subject languages are found: the first is characterized as consistent null subject languages and corresponds to grammars such as Italian and Spanish; while this type admits 3rd person singular null subjects receiving a definite interpretation, it does not have null indefinite pronouns (corresponding to a null “one”), as shown in the following Italian sentence (Holmberg 2010a, p. 91):

(8) Gianni<sub>1</sub> dice che (\*lui<sub>1</sub>) vuole comprare una macchina.

Gianni says that he wants buy a car

The second variety is characterized as partial null subject languages and corresponds to grammars such as Brazilian Portuguese (BP), Finnish or Marathi; this type admits null 3rd person singular indefinite or arbitrary subject pronouns, as respectively shown in (9a) and (9b), although definite pronouns can optionally be null only when they are locally c-commanded by an antecedent, as shown by the difference in grammaticality between (9c) and (9d) (Holmberg 2010a, pp. 91-93):

(9) a. É assim que faz o doce. (BP)

is thus that makes the sweet  
“This is how one makes the dessert.”

b. Tässä tuolissa istuu mukavasti. (Finnish)

this-IN chair-IN sits comfortably  
“One can sit comfortably in this chair.”

c. Ram<sub>1</sub> mhanala ki (tyani<sub>1</sub>) ghar ghetla. (Marathi)

Ram said that he house bought  
“Ram said that he bought a house.”

d. Juha<sub>1</sub>ei ole sanonut mitään, mutta Pauli<sub>2</sub> sanoo että \*Ø<sub>1</sub> haluaa (Finnish)

ostaa uuden auton.  
“Juha<sub>1</sub> hasn’t said anything, but Pauli<sub>2</sub> says that he<sub>1</sub> wants to buy a new car.”

In order to account for the distributional patterns shown above, Holmberg (2005, 2010a) proposes that «the property which consistent NSLs have, that partial NSLs do not

have, is a D(efinite)-feature as part of the  $\phi$ -feature make-up of finite T», together with the two following assumptions (Holmberg 2010a, p. 94):

- (10) a. Pronouns are either DPs, with the structure [<sub>DP</sub> D [ <sub>$\phi$ P</sub>  $\phi$  [<sub>NP</sub> N]]], or  $\phi$ Ps;  
 b. Null pronouns are  $\phi$ Ps.

According to this proposal, in consistent null subject languages «when T probes a  $\phi$ P subject, and has its unvalued  $\phi$ -features valued by the subject, the resulting union of the  $\phi$ -features of T and the subject yields a definite pronoun» (Holmberg 2010a, pp. 94-95). This would account for the impossibility of having null indefinite pronouns in finite, active clauses (as shown in (8)), as the D feature in T is inherently definite. Conversely, in partial null subject languages, what allows the occurrence of null indefinite subject pronouns (as shown in (9a-b)) and, at the same time, forbids definite null subjects unless they are referentially dependent on some antecedent (as shown in (9c-d)) is that «the probe-goal relation between T and a null  $\phi$ P subject does not supply a definiteness value», the result of this agreement relation being «a D-less, thus indefinite, subject pronoun» (Holmberg 2010a, p. 95). Crucially, however, Holmberg (2010a) observes that treating this D-feature in finite T as valued gives rise to two problems concerning consistent null subject languages. First, the well attested possibility for some of these languages to have indefinite lexical subjects would be ruled out *a priori*. Second, this same scenario would fail to capture the fact that, as noted by Samek-Lodovici (1996) and shown in the following Italian example, «null subjects, particularly 3rd person null subjects, are dependent on an antecedent in consistent NSLs, too» (Holmberg 2010a, pp. 95-96):

- (11) a. Questa mattina, la mostra      è stata visitata da Gianni. Più tardi \*Ø/egli/lui  
           ha visitato l'università. (Samek-Lodovici 1996)  
           This morning    the exhibition was visited      by Gianni. Later      he/he  
           visited the university  
           “This morning the exhibition was visited by Gianni. Later he  
           visited the university.”  
 b. Questa mattina, Gianni ha visitato la mostra.      Più tardi Ø ha visitato  
           l'università.  
           This morning    Gianni visited      the exhibition. Later      visited  
           the university  
           “This morning Gianni visited the exhibition. Later he visited the university.”

Besides the fact that in (11b) the possibility of having a null subject is assumed to depend on a pronoun-antecedent relation rather than on the specification of a valued D-feature in finite T, the other aspect emphasized by Holmberg is that there seems to be particular conditions on this relation. More precisely, as shown by the impossibility of having a null subject in (11a), the condition that applies here is that, in Frascarelli's (2007) terms, «the interpretation of a referential *pro* depends on a matching relation with a specific type of Topic: the so-called 'Aboutness-shift' Topic» which is always located in the C-domain, either overtly or covertly (Frascarelli 2007, p. 693):

While every Topic expresses, in some sense, pragmatic 'aboutness' (Reinhart 1981), the specificity of the Aboutness-shift Topic is to newly propose or reintroduce a topic in the discourse (cf. Givón 1983). (Frascarelli 2007, p. 697)

According to Holmberg, in contexts such as (11b) there is a null Aboutness-shift topic mediating the co-indexing relation between the null subject and its antecedent, which in turn corresponds to the Aboutness-shift Topic of a preceding clause, as represented in (12) (Holmberg 2010a, p. 96):

- (12) [CP <Gianni<sub>1</sub>> [questa mattina Gianni<sub>1</sub> ha visitato la mostra]]  
 [CP <Ø<sub>2</sub>> [più tardi ha  $\phi$ P<sub>2</sub> visitato l'università]]  
 1 = 2

In fact, *contra* Frascarelli (2007), Holmberg argues that «the referential index of the null subject then ultimately comes from [...] the index of a spelled out DP in the preceding discourse, via a chain of A-topics» (*ibidem*). Consequently, what rules out the possibility of having a null subject pronoun in (11a) is that an overt pronoun like *egli* or *lui* proposes the new topic *Gianni* in lieu of the topic of the preceding sentence, while a null pronoun would necessarily refer to *la mostra*. On the other hand, the null pronoun in (11b) is still dependent on an Aboutness-shift topic, the only differences being that, in this case, the latter is not overt and reintroduces a topic already present in the discourse (*Gianni*):

[...] following Frascarelli (2007), I assume: (a) that an Aboutness-shift topic (henceforth A-topic) is always syntactically represented in a designated A-topic position in the articulated C-domain, either overtly (for instance in the Italian clitic left-dislocation construction) or covertly; and (b) that the antecedent of a null subject is a null A-topic base-generated in the C-domain of the clause immediately containing the null subject. This null A-topic is a copy of

an A-topic, which may or may not be null, in the locally preceding discourse (Holmberg 2010a, p. 96)

Returning to Holmberg's (2010a) proposed distinction between consistent null subject languages and partial null subject languages, what the A-topic does in languages of the former type is value the unvalued D-feature (labelled [uD]) hosted by finite T as well as checking the EPP feature in T. In these terms, the reason why consistent NSLs cannot have a null indefinite subject is that, since this element is «a bare  $\phi$ P which cannot value [uD]», [uD] in T cannot be valued and, therefore, the derivation crashes (Holmberg 2010a, p. 97). Conversely, as far as partial null subject languages are concerned, «in a language without uD in T such a pronoun can only be interpreted as impersonal, that is either as generic (inclusive or exclusive) or non-thematic» (*ibidem*).

If on the one hand Holmberg (2010a) delves into the typological properties distinguishing the types of null subject languages from one another, it is actually Roberts (2010) which introduces «the technical ideas which [...] motivate and restrict the environments of pronoun-deletion which give rise to null subjects» (Roberts 2010, p. 63). Building on his analysis of Romance clitics and cliticization, Roberts proposes an account of the null subject phenomenon in terms of incorporation, here meant as «head-movement, adjoining a minimal category to a minimal head» (Roberts 2010, p. 64). Assuming the existence of an Agree relation in the sense of Chomsky (2001b) between finite T (which acts as a Probe) and the subject pronoun (which acts as a Goal) whereby «the valuing operation consists of copying the values of the valued counterparts of the features into the blank value matrices of the unvalued features», Roberts's core idea is that these two agreement-related elements may form a chain (Roberts 2010, p. 60). This happens depending on two conditions: that finite T has unvalued  $\phi$ -features, and therefore effectively probes for a goal which can value them, and that the features of the subject pronoun represent a subset of the features of finite T. This latter requisite is referred to as *defectivity* and it is, according to Roberts, the same general relation occurring between the incorporated clitic's  $\phi$ -features and  $v^*$ 's unvalued  $\phi$ -features in Romance languages (cf. Roberts 2010, pp. 64-65):

Incorporation can take place only where the features of the incorporee are properly included in those of the incorporation host. (Roberts 2010, p. 65)

Besides the central role of Agree, Roberts's (2010) account of subject-deletion in null subject languages also relies on the fact that, in order to form a chain with finite T, the pronoun must not have a label distinct from its probe (cf. Roberts 2010, p. 64). As seen above in (10), this is precisely what Holmberg (2005, 2010a) proposes with respect to null pronouns, which are regarded as acategorical  $\phi$ Ps (although Roberts (2010) regards *pro* as a D (see Roberts 2010, p. 73), this and other differences are not relevant to the scope of the present thesis). Given these conditions, incorporation of the subject pronoun in finite T occurs when «copying the features of the defective goal exhausts the content of the goal», that is, precisely in the case when «the operation is not distinguishable from the copying involved in movement» (Roberts 2010, p. 66). As summed up by Holmberg:

This is what Roberts (in press, 2009a [=2010, A.R.]) refers to as *incorporation* of the pronoun in T. The distribution of features is the same as if the subject had undergone head-movement, incorporating in T, but no movement has taken place, only Agree. The result is, however, that the subject is formally a *copy* of T. Still following Roberts, this means that T and the subject form an argument chain, headed by T. (Holmberg 2010b, p. 21)

In these terms, *pro*'s silent nature can be derived in terms of *chain reduction*, that is, a PF-deletion process consisting in «the deletion of all identical copies in a dependency except the highest one (see Nunes 2004: 22f.)» (Roberts 2010, p. 66). Since in this probe-goal relation the highest copy is constituted by the features of T which have been valued by the subject  $\phi$ P, the latter is not pronounced and hence surfaces as a null subject.

Returning to Holmberg (2010a)'s analysis, in consistent null-subject languages the defective 3rd person pronoun's  $\phi$ -features are a proper subset of finite T's features as the latter element, after Agree, also has a D-feature valued by the A-topic in SpecCP and a tense feature. Thus, «the result is a definite null subject construction, with the referential index of the A-topic» (Holmberg 2010a, p. 98). In fact, as argued in Roberts (2010), the only case in which a subject cannot undergo incorporation/deletion is when this element is a lexical DP or a D pronoun, and this because «a lexical DP has a root, which is not copied by T under Agree» (*ibidem*):

The result is that the only pronouns that remain null are the ones that are linked to a null A-topic in SpecCP. The generalisations that 3rd person null subjects in consistent null-subject languages are always definite, and always refer to a person or object already introduced as a topic, are thereby explained. (Holmberg 2010a, p. 99)

As far as partial null subject languages are concerned, their proposed lack of a uD feature in finite T does not in principle hinder the subject form being incorporated in T. Therefore, also their subjects pronouns can undergo the same deletion mechanism which is at work in consistent null-subject languages:

In [...] partial null-subject languages, the subject can still be null essentially by the same derivation as in the consistent null-subject languages: T probes for  $\phi$ -feature values. The subject's  $\phi$ -feature values are copied by T, and the subject has its Case-feature valued in return. In the case where the subject is a bare  $\phi$ P, T will copy all the feature values of the subject. As a result T and  $\phi$ P form a chain, and the subject remains null, by chain reduction. (Holmberg 2010a, p. 101)

However, what does change partial null subject languages with respect to consistent null-subject languages is the interpretation of the null  $\phi$ P which, in the absence of uD in T, «cannot be that of a definite pronoun» (*ibidem*). Concerning the availability of antecedent-controlled null definite subjects in Marathi, BP and Finnish (as shown in (9c)), according to Holmberg these covert elements are different from their indefinite counterparts in that they are additionally specified with an unvalued D-feature. This proposal makes two interdependent predictions. The first is that, in partial null-subject languages, definite null subjects do not incorporate in T and, consequently, are free to move to SpecTP in order to check the EPP. Secondly, starting from the assumption that «the indefinite, generic null subject does not have an A-topic antecedent» by which the EPP can be checked, in these sentences the EPP condition has to be satisfied by some other category as the null indefinite pronoun is incorporated in T (Holmberg 2010a, p. 102). This is exactly what happens in the Finnish example (13) (*ibidem*):

- (13) a. Jari sanoo että tässä istuu mukavasti.  
           Jari says   that here sits   comfortably  
           “Jari says that one can sit comfortably here.”  
           ≠ “Jari says that he sits comfortably here.”  
       b. Jari sanoo että Ø istuu mukavasti   tässä.  
           Jari says   that   sits   comfortably here  
           “Jari says that he sits comfortably here.”  
           ≠ “Jari says that one can sit comfortably here.”

As noted by Holmberg, while in (13a) the EPP condition is satisfied by the place adverb *tässä*, which has undergone movement to SpecTP instead of the indefinite null subject, in (13b) it is the subject pronoun that has moved to Spec-TP satisfying the EPP (cf. Holmberg 2010a, p. 102):

It appears, then, that while definite null subjects in consistent NSLs are incorporated  $\phi$ Ps which are interpreted as definite by virtue of a valued D-feature in T, definite null subjects in BP, Finnish, and, I assume, Marathi are DPs which have been second-merged with SpecTP. (Holmberg 2010a, p. 103)

According to Holmberg, in this position the null subject receives a definite interpretation by virtue of being controlled by a local antecedent in a higher clause, which also contributes to its licensing as a null element «by virtue of an extended version of chain reduction» (Holmberg 2010a, p. 104). Crucially, in this way the condition stating that in partial null-subject languages definite null subjects must be bound by a higher argument also directly follows. Summing up, what emerges from Holmberg's analysis of null-subject languages is that, contrary to what might be expected, there are actually two different types of null subject pronouns: one is the D-less  $\phi$ P characteristic of null subject languages in general which, due to its defective nature, undergoes incorporation in T and cannot satisfy the EPP. The other one, which is present only in a subset of these languages (the partial null-subject languages), has an unvalued D-feature which, while requiring to be valued by an argument in a higher clause, prevents incorporation in T and allows the pronoun to move to SpecTP and check the EPP.

In conclusion, the analyses put forth in Roberts (2010) and Holmberg (2010a) can arguably be regarded as a notable joint attempt to actualize the null-subject parameter by explaining the distribution of pro-drop across null-subject languages in terms of richness of  $u\phi$  and  $uD$ -feature content in T. Crucially, what is particularly important about these proposals is that the patterns of cross-linguistic variation upheld in them not only do conform to Robert & Holmberg's (2010) hierarchical model of parametric variation but, especially assuming Holmberg's account, «are also plausibly derived from more primitive conditions on spell-out and linearisation» which are highly desirable from a strong minimalist perspective (Holmberg 2010a, p. 98, n. 7).



#### 4.2.2 – The V-to-T movement parameter

Among the parameters that retained their theoretical status in current generative theory, a notable case is the V-to-I(nfl) movement parameter (now commonly referred to as V-to-T(ense) from Chomsky (1995a) onwards) as reformulated in Biberauer & Roberts (2010, 2012) and evaluated typologically in Holmberg & Roberts (2013). As Biberauer & Roberts (2012) introduces some important developments on the nature of verb movement with respect to both Biberauer & Roberts (2010) and Holmberg & Roberts (2013), this work will here be regarded as a state of the art account for this phenomenon and, therefore, it will be reviewed after the other two papers referred to here despite its publication date.

Starting from Biberauer & Roberts (2010), this paper begins by briefly examining the hypothesis according to which both the licensing of null-subjects and V-to-T movement directly depend on the morphological richness of verbal inflection. While the idea of a correlation between rich agreement and pro-drop dates back to Taraldsen's (1978) «intuitive idea [...] that where there is overt agreement, the subject can be dropped» (Chomsky 1993 [1981a], p. 241), the hypothesis that rich agreement leads to both V-to-T movement and pro drop has been more lately put forth by Koenenman & Neeleman (2001), which also proposed a three-fold classification of languages according to their relative inflectional richness (cf. Biberauer & Roberts 2010, p. 264). This classification is then reformulated by Biberauer & Roberts in a typological fashion, as shown in (14) (*ibidem*):

- (14) a. the rich: V-to-T and null subjects (Italian, Greek, Spanish, etc.);  
b. the middle class: enough 'wealth' for V-to-T but not enough for null subjects (French, Middle English);  
c. the impoverished: neither V-to-T nor null subjects (Modern English, Mainland Scandinavian).

Although the idea of reducing V-to-T movement and the Null Subject Parameter to an individual inflectional property is certainly appealing, the most worrisome aspect of this proposal is that, as shown in the following examples, V-to-T movement is well attested also in non-null-subject languages. This is notably evident in French which, although exhibiting verb movement to T in tensed clauses on a par with other Romance languages such as Italian and Spanish (as shown in (15a)), has overt *there*-type expletives rather than null ones (as shown in (15b)) (Biberauer & Roberts 2010, pp. 264-265):

- (15) a. Jean embrasse souvent Marie.  
           John kisses often Mary  
           “John often kisses Mary.”  
       b. \*(Il) y a une licorne dans le jardin.  
           there is a unicorn in the garden  
           “There is a unicorn in the garden.”

Dismissing Koenenman & Neeleman’s idea that «poor agreement paradigms do not trigger [verb] movement or license null subjects, while rich agreement leads to both V to I and pro drop», not only do Biberauer & Roberts regard pro drop and V-to-T movement as two independently motivated syntactic properties but also make a clear distinction between the respective roles of agreement inflection and tense inflection in accounting for these phenomena (Koenenman & Neeleman 2001, p. 215). Starting from the assumption that rich person/number agreement inflection only motivates the licensing of null subjects (as argued in Roberts (2010) and Holmberg (2010a)) and is therefore not directly related to V-to-T movement, according to Biberauer & Roberts «the difference between Germanic and Romance» with respect to the latter syntactic property «is correlated with the richness of the inflectional (i.e. synthetic) marking of tense distinctions», here taken to be reflected by the number of distinctive oppositions within the tense-modality-aspect system of the language in question (Biberauer & Roberts 2010, p. 266):

Our approach thus postulates that there are two quite distinct types of ‘richness’ of verbal inflection: **agreement inflection** and **tense inflection**. We take tense inflection to include the marking of notional mood and aspect, as in the Romance subjunctive, futures, conditionals and imperfect forms. (*ibidem*)

Taking Germanic and Romance to prototypically represent non-verb-movement and verb-movement systems respectively, Biberauer & Roberts observe that, as far as tense inflection is concerned, «the Romance languages are considerably richer, in a clear intuitive sense, than the Germanic ones», as shown in (16-17) (*ibidem*):

(16) **Romance:**

French: *parle* (present indicative/subjunctive), *parlerai* (future), *parlerais* (conditional), *parlais* (imperfect), [*parlai* (preterite), *parlasse* (past subjunctive)]  
Italian: *parlo* (present), *parlerò* (future), *parlerei* (conditional), *parlavo* (imperfect), *parli* (present subjunctive), *parlassi* (past subjunctive), *parlai* (preterit)

Spanish: *hablo* (present), *hablaré* (future), *hablaría* (conditional), *hablaba* (imperfect), *hablé* (preterite), *hable* (present subjunctive), *hablase* (past subjunctive I), *hablara* (past subjunctive II)

(17) **Germanic:**

German: *spreche* (present indicative/subjunctive), *sprach* (past), *spräche* (past subjunctive)

English: *speak* (present), *spoke* (past)

Swedish: *snakker* (present), *snakket* (past)

The next step of Biberauer & Roberts's argumentation is to propose an alternative typology in which the "middle class" hypothesized by Koenenman & Neeleman (2001) is replaced with two distinct classes, which in turn correspond to (18b) and (18d), whose attributes reflect the possible discrepancy between the relative strength of agreement inflection on the one hand and tense inflection on the other. This new typology is represented below (Biberauer & Roberts 2010, p. 267):

- (18) a. Rich agreement and rich tense inflection: hence V-to-T and null subjects,  
e.g. Italian, Greek, Spanish, etc.
- b. Poor agreement but rich tense: hence V-to-T, but no null subjects,  
e.g. French, Middle English.
- c. Poor tense and poor agreement: hence no V-to-T and no null subjects,  
e.g. Modern English, Mainland Scandinavian.
- d. Rich agreement and poor tense: null subjects, but no V-to-T;  
no clear example.

After arguing that the properties triggering V-to-T movement should be kept distinct from the ones responsible for null subject patterns, Biberauer & Roberts present their own account of verb movement. Instead of resorting to a feature-strength-oriented model of syntax in the spirit of Chomsky's (1995a) checking theory but assuming, along current minimalist assumptions, that «in the Agree model [...] there is no longer any specific reason to assume a correlation between inflectional richness and the presence of a movement trigger», the authors adopt what they call a reprojection-based approach to account for the relation between rich tense and V-to-T movement (*ibidem*). A lexically-based approach in nature, Biberauer & Roberts's proposal is that what differentiates languages with a rich tense morphology from poorer systems in this respect is that in the former, but not in the latter, finite verbs are actually «compound elements consisting of V and a fully-specified T [...] formed in the Numeration» instead of categorically simple Vs (*ibidem*). Since such a finite verb has to merge with both a V-complement and a T-

complement in order to satisfy its (twofold) selectional requirements, according to Biberauer & Roberts V-to-T movement is a result of the fact that «the only way to satisfy the properties of the compound V+T element is to first merge it as V [...] and then to raise it, following the standard movement process, to project as what we usually think of as T» (Biberauer & Roberts 2010, pp. 267-268). In these terms, V-to-T movement is here held to be a consequence of the fact that the finite verb actually merges twice: under the V node forming a VP and, then, under the T node projecting a TP:

We are therefore proposing a form of partial reprojection, in that the T-features of the compound element determine the formation of the TP (the V-features do not, as these have played their role in forming the thematic domain of VP, although they must move with T as part of the compound V+T element). (Biberauer & Roberts 2010, p. 268)

The next part of Biberauer & Roberts (2010) is not basically concerned with providing explanation of the actual number of paradigmatic tense, modality and aspect distinctions involved in triggering verb movement, but rather devoted to the implications of their proposed typology of verb raising and null subject languages.

Turning now to Holmberg & Roberts's (2013) work, their analysis of the relationship between tense inflection and verb movement ideally begins where Biberauer & Roberts's (2010) one ended, namely, by extending Biberauer & Roberts's proposal about tense inflection and V-to-T movement to a broader set of languages. Holmberg & Roberts's starting assumption is given in (19) (Holmberg & Roberts 2013, p. 114):

(19) Move V to T if and only if T has rich inflection.

with rich inflection referring to «inflection for future tense and subjunctive or other moods» (*ibidem*).

The first language put to test in this respect is Chinese (Mandarin), a highly analytic system in which, as noted by the authors, «information regarding tense, mood and aspect is carried by separate “particles” of various kinds» (Holmberg & Roberts 2013, p. 112). As in the example (20) the elements in bold correspond to aspectual particles which are taken to be located in T (Holmberg & Roberts 2013, p. 114):

- (20) a. Ta **zai** dasheng chang ge. (Huang *et al.* 2009)  
           he PROG loud sing song  
           “He was singing loudly.”  
       b. Wo mei-**you** qiaoqiao de hui jia.  
           I not-PRF quiet DE return home  
           “I didn’t go home stealthily.”

the consequent prediction that V will not move but rather overtly occur adjacent to its object (or other verb complement) is therefore correct.

Omitting here any further mention to French, which has already been referred to above as fitting into the generalization in (19), another good piece of evidence for the relationship between rich tense inflection and verb raising is represented, according to Holmberg & Roberts, by those cases in which richness of inflection of T has been susceptible to diachronic reduction. Predicting that «as verbal inflection is eroded, then the position of the verb will change over time», the fact that languages such as English, Swedish and Haitian Creole have moved from an early stage featuring richer verbal inflection and V-to-T movement to a later stage in which both these properties have disappeared further supports the idea of a «correlation between the amount of verbal inflection a language has and the position of the verb in the sentence» (Holmberg & Roberts 2013, p. 115). As noted by the authors:

This is clearly seen in English: a set of verbal inflections (plural endings in the past and present tense) were lost in the 16th century, and, within at most a hundred years, the position of the verb had changed, i.e. V-to-T movement had been lost. The same can be observed in the history of Swedish, which moved from an “Icelandic-like” verbal inflection system with V-to-T movement to its modern system as described above during the Middle Ages (see Falk, 1993). Faroese may be just completing this set of changes [...]. Finally, the creolisation process that created Haitian Creole involved the stripping away of nearly all French inflection, and again we see that the verb does not move to T in the creole. (Holmberg & Roberts 2013, pp. 115-116)

The last pieces of typological evidence put forth in Holmberg & Roberts (2013) to corroborate the validity of (19) concern verb-initial languages, whose basic word order is either VSO or VOS. The first two examples are from Welsh, a VSO language, with the elements in bold representing T(ense) (Holmberg & Roberts 2013, p. 116):

- (21) a. Fe/mi **welais** i Megan.  
           PRT   saw     I Megan  
           “I saw Megan.”
- b. Fe/mi **wnes** i weld Megan.  
           PRT   did     I see   Megan  
           “I saw Megan.”

While (21a) represents the default word order for Welsh and all other Celtic languages, with the verb being not adjacent to the direct object as expected for a VSO system, (21b) shows an alternative AuxSVO construction, peculiar to Welsh, which involves the auxiliary *gwneud* (“do”) and a non-finite form of the verb. As «Welsh auxiliaries carry information regarding tense, aspect and mood», what is important here is that these elements can hence be located in T and, accordingly, (21b) can be attributed the following syntactic structure (with the root affirmative particle *fe/mi* being merged with TP in a higher position) (Holmberg & Roberts 2013, p. 116):

- (22) *fe/mi* [<sub>TP</sub> T [<sub>VP</sub> Subject [<sub>VP</sub> V Object]]]

What (22) shows is that in (21b) the subject enters the derivation by merging with the constituent formed by the verb and the direct object. This is in fact what is universally assumed for all natural languages, with the possible subsequent option of moving the subject out of the vP not occurring in Welsh. In these terms, while on the one hand (21b) shows that the non-finite verb stays *in situ* when clustering with an auxiliary, in (21a) what occurs instead is that «examples without auxiliaries and with VSO order involve V-to-T movement, giving [23]» (Holmberg & Roberts 2013, p. 117):

- (23) [<sub>TP</sub> V+T [<sub>VP</sub> Subject [<sub>VP</sub> Object]]]

Starting from the assumption that the word order Verb-Subject-Object involves verb-movement out of VP and over the subject, Holmberg & Roberts argue that, given the previously observed typological patterns, «Welsh and other VSO languages should have complex verbal inflection» (*ibidem*). In these terms, as noted by the authors:

Both Welsh and Irish certainly fit the bill, in that each has around twenty finite verb forms. Scots Gaelic and Breton may be more impoverished, but still significantly richer in this

respect than Modern English or Swedish. The principal VSO Semitic languages (Classical Arabic, Ge'ez and Biblical Hebrew), all have very rich verbal inflection, comparable to, say, Finnish. The same is true, as far as we are aware, of other Afroasiatic VSO languages such as Egyptian and Berber. (Holmberg & Roberts 2013, p. 117)

Turning now to other V-initial languages, the grammatical system chosen by the authors in order to further test the validity of (19) is a Polynesian language, Niuean. In this context, what differentiates this V-initial language from VSO systems like Welsh is that, as observed by Massam (2000, 2005), Niuean has a predicate fronting rule that moves the maximal projection VP to a position within TP, thus yielding VOS structures such as (24) (*ibidem*):

(24) [TP [VP V O] T [VP S .. (VP)]] – VOS]

In addition to (24), in which the NP object «does not move out of VP to a checking position, but instead remains *in situ* within the VP, and gets fronted along with the verb by predicate-fronting», this same predicate-fronting rule is also responsible for the derivation of VSO structures (Massam 2000, pp. 106-107). As shown in (25) (Holmberg & Roberts 2013, p. 117):

(25) [TP [VP V (O)] T [VP S [O (VP)]]] – VSO

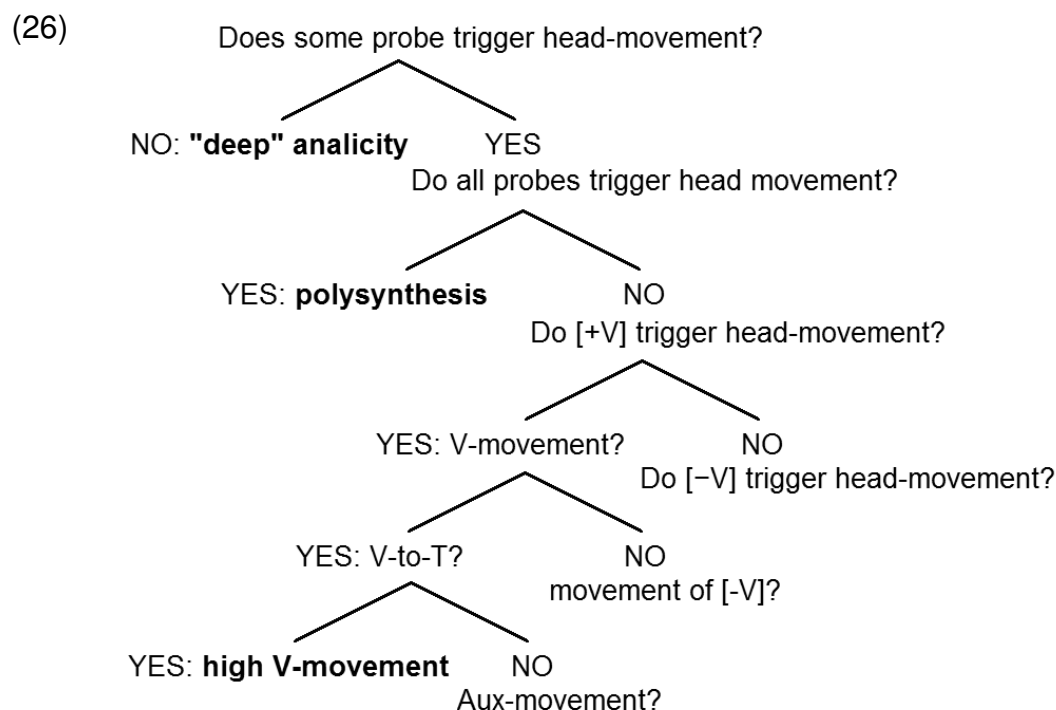
this happens precisely when, as proposed by Massam, «the object has undergone object shift out of VP prior to the fronting of the VP» (Massam 2005, p. 228). Assuming that Massam's account is correct not only for Niuean, but for all other verb-initial languages which exhibit both VSO and VOS orders, what necessarily follows from this whole analysis is that VOS/VSO languages of the Niuean type should have impoverished verb morphology (cf. Holmberg & Roberts 2013, p. 117). According to the authors, this property – and hence the lack of verb movement – should in fact follow from the incompatibility between V-to-T movement and the VP predicate-fronting rule characterizing this typological group:

In terms of our proposal that verb-movement to T is necessarily connected to richness of verbal inflection, we arrive at the prediction that VOS/VSO languages of the Niuean type will have impoverished verb morphology. If these languages had rich inflection on verbs, then this would have to give rise to V-to-T movement, an option incompatible with the VP-

movement containing an unmoved verb that Massam proposes (cf. also Massam, 2005:239).  
(Holmberg & Roberts 2013, p. 117)

The argument is compelling, and Holmberg & Roberts's conclusion is correct; not only does Niuean not have rich verb inflection, with tense/aspect information being entirely carried by sentence-initial particles, but this is also the case for many other VSO-VOS Oceanic languages which either lack tense and agreement inflection (as Samoan, Maori, Rapanui, Tongan and Chamorro) or have a tense prefix but no inflection for agreement (like Malagasy) (cf. *ibidem*).

Finally, regarding the status of the connection between V-to-T movement and rich tense morphology in a parametric model of linguistic variation, Biberauer & Roberts (2012) identify the generalization in (19) with a parametric option belonging, from a broader perspective, to the parametric hierarchy determining word structure, here meant as spanning both maximally analytic and polysynthetic systems (the latter of which will be discussed later in this chapter) with all the in-between options. In the spirit of Roberts & Holmberg's (2010) emergent conception of parametric variation, this hierarchy is represented in (26) (adapted from Biberauer & Roberts 2012, pp. 276, 281):



Although in this parametric schema high V-movement languages correspond to one of the two bottom branches, neither one of these two options is actually the most marked in an absolute sense, as the binary choice corresponding to Aux-movement further



extends its scope downwards by progressively embracing more and more marked options (cf. Biberauer & Roberts 2012, p. 281). Similarly to what has been pointed out regarding the Null Subject Parameter, in this respect the «V-to-T» parameter» of Biberauer & Roberts (2012) is conceived as a mesoparameter, as the extra feature triggering head-movement is here taken to be specified on all the heads of the natural class [+V] (Biberauer & Roberts 2012, p. 271). More precisely, assuming Roberts's (2010) account of head-movement in terms of incorporation, according to Biberauer & Roberts the feature specification triggering V-to-T corresponds to the verb's categorial label V (which is probed by the unvalued V-feature in T) and, following Biberauer & Roberts's (2010) observations about verb movement and the richness tense morphology, a set of unvalued features regarding the language's tense-aspect system:

Here we propose that lexical verbs in systematic verb-raising systems [...] differ from those in consistent non-raising systems [...] in that they enter the syntactic derivation in already-“verbalised” form, i.e. as heads specified [V]. This verbal specification is necessary to allow verbs to bear a formal feature for which higher heads can probe (see Biberauer & Roberts 2010) and also to allow the verb to bear unvalued V-related formal features (e.g. [Asp:\_\_\_] and [T:\_\_\_]) for which higher heads can probe [...]. (Biberauer & Roberts 2012, p. 274)

Crucially, the fact that in this later reformulation V-to-T movement is accounted for not in terms of reprojection but rather in terms of «an Agree relation involving a defective goal» suggests that, differently from Biberauer & Roberts's (2010) original proposal, verb movement is regarded as an externalization-related phenomenon in the same manner as pro drop rather than the product of narrow syntax only (Biberauer & Roberts 2012, p. 273). In these terms, although Biberauer & Roberts (2012) does not delve further into this particular issue, this conclusion nonetheless serves us as a confirmation that the V-to-T movement parameter not only conforms to the hierarchical/emergent model of Roberts & Holmberg (2010) but also, more in general, to the minimalist conception of overt movement in terms of Internal Merge plus deletion of the lower copy.

#### 4.2.3 – The V-to-C movement parameter

Compared to the rest of the GB-parameters reviewed in this chapter, the V-to-C parameter is arguably the one whose latest formulation most closely resembles the one put forth in its original proposal. As it will be pointed out below, however, according to more modern parametric approaches, the complementary distribution of finite verb and lexical complementizer in verb-second languages discussed in Den Besten (1983) is no more attributed to the effect of an inherent property of the verb, but rather to that of an independently postulated feature requiring the complementizer to be overtly realized at PF. This hypothesis is proposed in Roberts (2004).

In this paper, which gives a comparison of the Celtic preverbal particle system with the V2 system of Germanic languages, Roberts begins his treatment by analyzing which position these elements take at a syntactic level. For the sake of simplification, only the root affirmative particle *fe*, *mi* and *y* are taken into account here (Roberts 2004, p. 298):

- (27) a. Root affirmative *fe*, *mi*, *y*.

Fe/mi welais i John.

PRT saw I John.

“I saw John.”

- b. Direct relative:

y dynion a ddarllenodd y llyfr.

the men PRT read-3sg the book

“the men who read the book.”

- c. Indirect relative:

y dynion y dywedodd Wyn y byddant yn darllen y llyfr.

the men PRT said Wyn that will-be-3pL ASP read the book

“the men who Wyn said will read the book.”

Although «it seems clear that the particles just discussed occupy C», according to Roberts what is still not clear is the actual projection these elements occupy in a split-C system in the sense of Rizzi (1997) (Roberts 2004, p. 299). Rizzi’s overall system for the left periphery is given below (Rizzi 1997, p. 297):

- (28) ForceP > TopP\* > FocP > TopP\* > FinP

In this connection, Roberts builds on McCloskey (1996)'s observations about sentential adverbs in languages such as English and Irish. According to McCloskey, while on the one hand «adjunction to IP is suggested by the possibility of multiple attachment (with, of course, varying scopes) of members of the same class of adverbs», on the other hand «the crucial property of this class of adverbs [...] is that they may not adjoin to CP» (McCloskey 1996, p. 56). The effect of this general ban on the adjunction of adverbs to CP is shown in (29) (Roberts 2004, p. 299):

- (29) a. In general, he understands what's going on.  
b. It's probable that in general he understands what's going on.  
c.\*It's probable [<sub>CP</sub> in general [<sub>CP</sub> that he understands what's going on ]].  
d.\*[In general [that he understands what's going on]] is surprising.

As shown in (29c) and (29d), the readings which are ruled out in English are those according to which the sentential adverb would be interpreted as modifying the embedded *that*-clause. Assuming this condition to be universal, however, a problem arises when considering the behavior of sentential adverbs in Irish, as «the ordering-restrictions are, in essence, the reverse of those we have just observed», with the adverb apparently being adjoined to the left edge of an embedded clause (McCloskey 1996, p. 58). This is illustrated in (30) (Roberts 2004, p. 300):

- (30) Is dóíche [ faoi cheann cúpla lá [go bhféadfaí imeacht ]]  
 is probable at-the-end-of couple day that could leave  
 ADV C I

In these terms, McCloskey's idea was to postulate that the sentential adverb is actually adjoined to IP in both English and Irish but that, crucially, the systematic contrast between these two languages derives from the fact that Irish features a rule lowering C-to-I in finite clauses (cf. McCloskey 1996, p. 66):

On the view developed here, IP and CP are as distinct as they are in English or French. IP is therefore a legal adjunction site as in English, and C, as in English, reflects the selectional requirements of governing lexical heads. The difference is that these facts are obscured in Irish by the lowering of C to I<sup>0</sup>. (McCloskey 1996, p. 67)

Dismissing the idea of a C-to-I lowering rule and taking advantage of Rizzi's split-CP system, Roberts's (2004) alternative solution is that, although being both inserted into root Fin, complementizers like English *that* and Irish *go* are overtly realized on different heads (cf. Roberts 2004, p. 300). More precisely, according to Roberts «it may be that English Force overtly attracts Fin [...], while in Irish this is not the case», with sentential adverbs of the kind observed in (29-30) occupying an intermediate position between these two functional heads, as summarized in (31) (Roberts 2004, pp. 300-301):

(31) [ForceP [Force *that*] ... [TopP Adv ... [FinP [Fin *go*] IP]]]

Returning to Welsh, since in this language «adverbs [...] cannot intervene between the particle *y*, which introduces finite clauses, and the verb», as shown by the ungrammaticality of (32) (Roberts 2004, p. 301):

(32) \*Dywedodd ef *y* yfory bydd yn gadael. (Tallerman 1996)  
 said he PRT tomorrow he-will-be ASP leave

Roberts concludes that, similarly to what has been said in connection with Irish complementizer *go*, the Welsh particle *y* is in Fin as well. Crucially, this exact behaviour is extended by Roberts also to *fe/mi* which, as shown in the following examples, must analogously be adjacent to the finite verb with no adverb intervening between them (Roberts 2004, p. 298):

(33) a. Bore 'ma, *fe/mi* glywes i 'r newyddion ar y radio.  
 morning this PRT heard I the news on the radio  
 "This morning, I heard the news on the radio."  
 b. \**Fe/mi* bore 'ma glywes i 'r newyddion ar y radio.  
 PRT morning this heard I the news on the radio

After this introductory analysis, Roberts gets to the heart of the matter by comparing the Welsh particle system with the V2 system of Germanic languages. In this respect, one crucial aspect shared by Welsh *fe*, *mi* and *y* and German V2 concerns the fact that both of them characterize root affirmative clauses (Roberts 2004, p. 302):

- (34) a. Yesterday John danced.  
 b. Gestern hat Johann getanzt.  
 c. \*Gestern Johann hat getanzt.

Assuming that in V2 languages V is moved to Fin, according to Roberts «it is natural to see this V-movement as directly analogous to the merger of the root affirmative particles in Welsh» (Roberts 2004, p. 302). More precisely, the scenario considered by Roberts in this respect is that, in root affirmative contexts, Welsh particles *fe*, *mi* and *y* and German finite verb must be both phonologically realized in Fin in order to fulfil a common PF-realization requirement which, in turn, corresponds to the individual parametric property: «Fin requires a PF-realization when +finite» (Roberts 2004, p. 303):

If, following Roberts (2001), we consider parameterization to be a question of PF-realization (or not) of functional categories, with movement triggered just where the lexicon makes available no lexical item which can satisfy the realization requirement by merger, then we can understand this similarity as a manifestation of the fact that Welsh actually has the same parametric property as German: namely, that Fin must have a lexical realization at some stage in the derivation (Fin\* in the notation of Roberts 2001). (Roberts 2004, p. 302)

In this sense, the systematic difference between Welsh and a V2 language such as German derives from the fact that only the former system has root affirmative particles which can be merged in Fin\*, the latter being forced to resort to V-to-C movement in order to satisfy this same requirement (cf. *ibidem*). By the same reasoning, the root-embedded asymmetry in V-to-C movement displayed by German can be accounted for by the availability, in this language, of non-root complementizers which, by being merged into embedded Fin, prevent movement of the finite verb only in embedded contexts. While the idea of a complementary distribution of preposed finite verbs and lexical complementizers was introduced by Den Besten (1983), Roberts's full explanation is actually more complex as it also involves a specific property of all Germanic languages, namely, that Force attracts a Fin complementizer in embedded declaratives – probably in order to satisfy a lexicalization requirement in much the same way as it has been argued above for Fin\* (cf. Roberts 2004, p. 303):

Let us suppose, then, that a selected Force position has features in virtue precisely of being selected by a higher predicate. The fact that typical complementizers like English *that* raise from Fin to Force must then be attributed to the fact that selected Force triggers overt Fin-

movement (presumably because it requires PF-realization). Let us suppose that the verb-movement part of V2 is a reflex of the fact that Fin requires a PF-realization, so in this case we have Fin\*. In these terms, we can see that complementizers are able to satisfy this requirement in embedded clauses, even if they subsequently raise to Force. So all the Germanic languages, including English, have Fin-to-Force movement where Force is selected (i.e., in embedded clauses). (Roberts 2004, p. 303)

According to Roberts, this property is the cause not only of the fact that English *that* is overtly realized higher in the left periphery than Irish *go* (as shown in (31)) but also of the fact that, as shown in (35), «German [...] requires embedded V2 exactly where the complementizer is missing» (*ibidem*):

- (35) Ich glaube, *gestern* *habe* Maria dieses Buch gelesen.  
 I believe yesterday has Maria this book read  
 “I believe Maria read this book yesterday.”

Summing up, while in both Welsh and German nonselected/declarative Fin requires to be overtly realized at PF, only the former language can fulfil the effects of «the Fin\* parameter» in root clauses by means of merging *fe/mi* or *y* into Fin\*, with German having to resort to verb movement instead due to its unavailability of Celtic-style root particles (Roberts 2004, p. 318).

Finally, there is a consideration of what might be the format of Roberts’s (2004) V-to-C parameter. Up to this point, this parameter has not been linked to any parametric hierarchy in light of Roberts & Holmberg’s (2010) proposed model for linguistic variation. Crucially, however, the reason for this could be due to the nanoparametric nature of this phenomenon. In other words, the fact that V-to-C movement is dependent on a feature localized to a specific head (Fin) not only characterizes it as a lexical idiosyncrasy, but also hinders its placement into a hypothetical parametric hierarchy.

In conclusion, the account of V-to-C movement proposed in Roberts (2004) represents a development of Den Besten’s (1983) original observations regarding V2 languages. While still building on the idea that presence of an overt complementizer blocks V-to-C movement, Roberts adapts this generalization to the split-C system of Rizzi (1997) by both explicitly arguing for the parametric status of the feature specification of Fin in V2 systems and attributing the latter’s characteristic root-embedded asymmetries in verb movement to the absence of root affirmative particles.

#### 4.2.4 – *The polysynthesis parameter*

Regarding the advancements of parametric theory with respect to the nature of syntactic incorporation, Baker's book *The Polysynthesis Parameter* (1996) ideally represents the direct development of the theory of incorporation which was put forth in Baker (1988). While Baker's account is explicitly macroparametric and, therefore, incompatible with Minimalism, some cues from his analysis of incorporation and other key syntactic properties characterizing polysynthetic languages have been lately reinterpreted along the lines of Roberts and Holmberg's (2010) parametric model by Biberauer *et al.* (2014).

Contrary to the main trends in Minimalism and in line with his own strenuous defence of the traditional macroparametric approach to typological variation, the analysis of polysynthetic languages carried out by Baker (1996) revolves around the idea that the difference between polysynthetic languages such as Mohawk and the other language types is grammatically relevant, and accountable for by the following parameter:

(36) The Polysynthesis Parameter (informal):

Every argument of a head element must be related to a morpheme in the word containing that head.

(Baker 1996, p. 14)

As the set of morphemes which Baker refers to in (37) corresponds to agreement morphemes and incorporated roots (cf. Baker 1996, p. 15), the Polysynthesis Parameter can in turn be formulated, in more formal terms, as the Morphological Visibility Condition (MVC), here meant as a parametrized constraint on  $\theta$ -role assignment which affects Mohawk and typologically similar languages:

(37) The Morphological Visibility Condition (MVC):

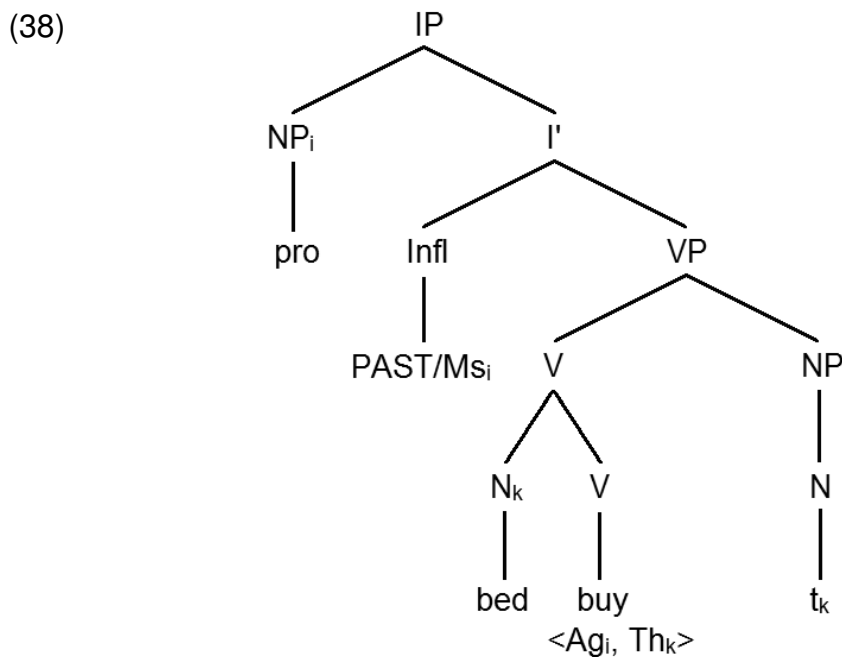
A phrase X is visible for  $\theta$ -role assignment from a head Y only if it is coindexed with a morpheme in the word containing Y via:

- (i) an agreement relationship, or
- (ii) a movement relationship

(Baker 1996, p. 17)

As far as polysynthetic languages are concerned, what the MVC states is that syntactic arguments cannot each be assigned a thematic role by the verb directly but, as a

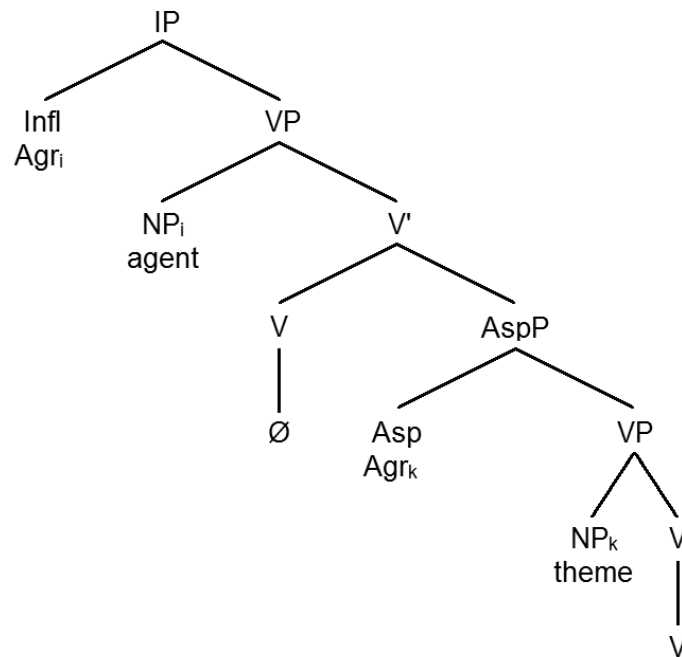
preliminary condition, they must first be made visible for theta-role assignment by the presence, on that verb, of either an agreement morpheme or an incorporated root. While building on the observation that, in polysynthetic languages such as Mohawk, «full NPs, when they appear, have the status of some kind of adjunct or modifier», according to the MVC the Theta Criterion has still to be satisfied syntactically (Baker 1996, p. 11). In this respect, Baker's hypothesis still relies on the idea that «the morphemes on the verb do not *replace* conventional argument phrases», as implied by the Uniformity of Theta Assignment Hypothesis (UTAH) originally put forward in Baker (1988), but with a different twist (Baker 1996, p. 15). More precisely, according to Baker (1996), NP arguments still exist in polysynthetic languages, but are actually *pro* elements. This is shown in (38) (Baker 1996, p. 16):



Regarding the options required to satisfy the positive setting of the MVC, both the movement relationship and the agreement relationship referred to in (37) are assumed by Baker to be the product of the same syntactic operation. In the first case, lexical roots are incorporated essentially in the same way to what had been previously proposed in Baker (1988), that is, as a result of verb movement. Similarly, in the second case, the characteristic subject and object agreement which can be found on the verb in polysynthetic systems results from head movement of V to Infl and Asp respectively, these being the positions in which agreement morphology is base-generated according to Baker's proposed structure of transitive clauses in Mohawk (Baker 1996, p. 231):



(39)



In these terms, although «no part of the Polysynthesis Parameter is really new» with respect to «the fact that lexical roots and affixes can represent arguments», the new aspects introduced by Baker (1996) of interest to the present thesis concern the cluster of properties which would derive not only *directly* from the positive setting of the Polysynthesis Parameter, but also *indirectly* from the interaction of this latter's effects with some other properties characterizing polysynthetic languages (Baker 1996, p. 19). In line with the proposed macroparametric status of the MVC, which is argued to be a global grammatical property rather than a lexical, microparametric one, according to Baker all polysynthetic languages must display a systematic set of syntactic properties which are in turn similar to those which Hale (1983) argued to be the classical features characterizing non-configurational languages (with the only exception of discontinuous nominal expressions, which are actually more restricted in polysynthetic systems than in non-configurational ones):

*Strikingly, every one of these languages has full and obligatory agreement paradigms for both subject and object (allowing for the possibility of some phonologically null third person forms). Not surprisingly, the languages also allow argument-drop and at least some degree of freedom in word order. This overall pattern is highly significant. There seems to be an implicational universal: all languages with full-fledged noun incorporation phenomena fall within the class of nonconfigurational head-marking languages. [...] However, the Polysynthesis Parameter explains why the two properties are related. This set of languages I will henceforth refer to as the true polysynthetic languages, and I will draw on them for comparison purposes throughout this book. (Baker 1996, p. 20)*

At this point, the syntactic properties dealt with by Baker are the possibility of argument drop and free word order. The former is shown in the following example from Mohawk, in which both the subject NP and the object NP can be omitted (Baker 1996, p. 41):

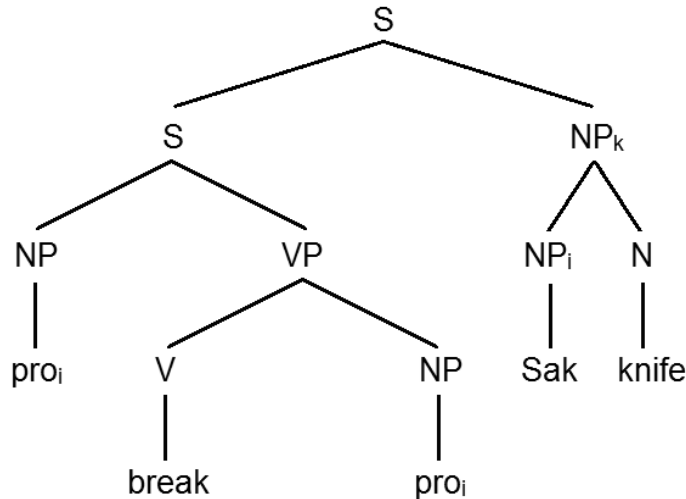
- (40) a. Wa'-ke-tshΛ'ri-'.  
 FACT-1sS-find-PUNC  
 "I found it."  
 b. Wa'-ke-tshΛ'ri-'                      kíkΛ    káhure'.  
 FACT-1sS-find-PUNC    this    gun  
 "I found this gun."

As far as (40) is concerned, the optional status of full nominals in Mohawk directly derives from the fact that, as anticipated above, «the MVC forces all overt NPs to be adjuncts» while all verbal arguments are null (Baker 1996, p. 42). An interesting example of this is represented by the possibility, in Mohawk, of co-indexing the possessor nominal of the overt object NP with the subject pronoun associated with the complex verb (as shown in (41)), with the same possibility being ruled out in English by Condition C (as shown in (42)) (Baker 1996, pp. 43, 45):

- (41) Wa'-t-há-ya'k-e'                      Sak raó-[a]'share'.  
 FACT-DUP-1sS-break-PUNC    Sak MsP-knife  
 "He broke Sak's knife." (coreference OK)  
 (42) \*He broke John's knife (in a fit of rage).

Taking examples such as (41) as evidence that the subject does not c-command the overt object, Baker's account of polysynthesis therefore assumes that overt NPs must be base-generated in adjoined positions, with the Projection Principle being satisfied by a series of null pronouns appearing in each available argument position. The structure of Mohawk sentence (41) thus corresponds to (43), with the overt NP object adjoined to the S node and thus outside the c-command domain of the covert null NP subject argument (Baker 1996, p. 47):

(43)



As noted by Baker, the adjunct status of overt NPs determines not only their optionality, but also their possibility for occurring on either side of the clause, and all of this «without changing the basic hierarchical properties of the structure» as «the subject pronoun does not c-command into the understood object, whichever side that NP appears on» (Baker 1996, p. 48). This freedom of order is evident when comparing Mohawk sentence (40b) with its equivalent (44) (Baker 1996, p. 41):

- (44) KíkΛ káhure' wa'-ke-tshΛ'ri-'.  
this gun FACT-1sS-find-PUNC

Although the properties of argument drop and variable word order both depend on the fact that, in Mohawk, overt NPs are not allowed to appear in argument position, the very reason for this latter requirement lies, according to Baker's account, in the relationship between the MVC and a rule, specific to polysynthetic systems, whereby heads cannot assign Case to argument positions, but instead only to the agreement morphemes which obligatorily appear on the verb. This rule is formulated in (45):

- (45) An agreement morpheme adjoined to a head X receives that head's Case at S-structure/PF.  
(Baker 1996, p. 86)

In these terms, the reason NPs are not allowed to appear in argument positions in Mohawk is because they would be ruled out by the Case Filter, which Baker reformulates as follows:

(46) The Case Filter:

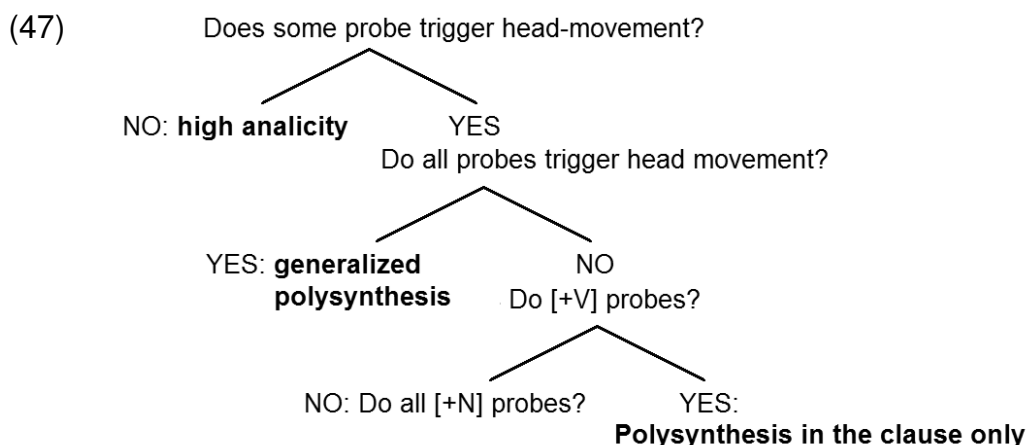
\*NP without Case if NP has phonetic features and is in an argument position.

(Baker 1996, p. 84)

Since according to (46) null NPs do not have to have Case, all available A-positions cannot be occupied by overt NPs, which in turn can therefore only appear as adjuncts in order to satisfy Baker's Case Filter. Summing up so far, what emerges from Baker's account is that polysynthetic languages differ from non-polysynthetic ones not only in the setting of the Polysynthesis Parameter but also in the properties attributed to their respective A-positions, which are in turn defined by virtue of some other independent grammatical properties.

Although Baker's (1996) work represents one of the most, if not the most, notable investigations into the nature of polysynthesis from a P&P perspective, both the format of his proposed "Polysynthesis Parameter" and the overall theoretical machinery he resorted to in order to account for what he holds to be the defining properties of polysynthetic languages appear, given its background and the time of writing, rather outdated. However, assuming that the incorporation-related phenomena found in this linguistic type are not essentially distinct from the instances of head-movement which are held to be at the basis of V-to-T and pro-drop, it seems promising to account for the possibility of argument drop and free word order featured by languages such as Mohawk within the parametric model formulated in Roberts & Holmberg (2010).

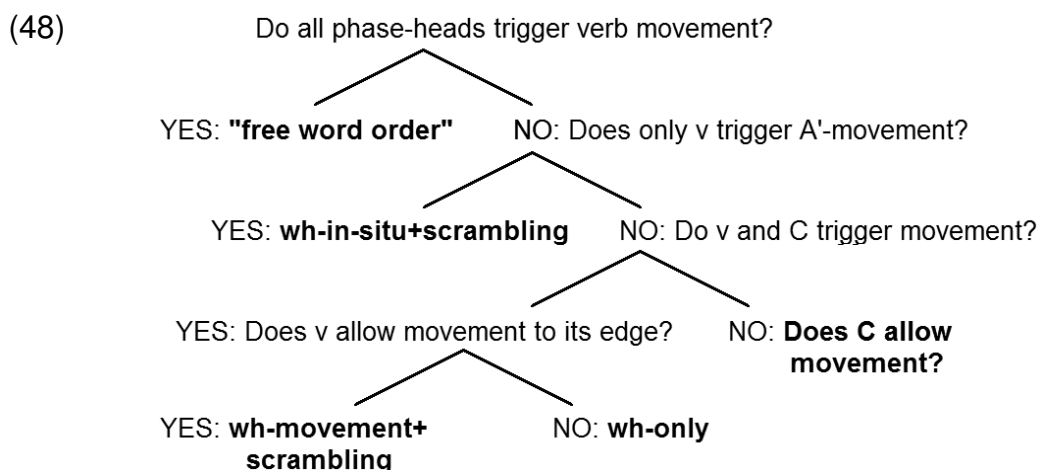
As briefly mentioned in connection with V-to-T movement, the incorporation processes which are the distinctive characteristic of the polysynthetic languages investigated by Baker (1988, 1996) are regarded as the second macroparametric option emerging from the hierarchy determining word structure (Biberauer *et al.* 2014, p. 114):



Assuming with Baker (1996) that incorporation essentially corresponds to head-movement affecting all lexical categories, according to Biberauer *et al.* (2014) such a conception of polysynthesis fits nicely into Roberts's (2010) account of head-movement in terms of Agree between a probe and a defective goal. In this light, polysynthetic systems emerge when the feature inducing incorporation is generalized on all functional heads triggering overt movement, with the only other less marked option available being represented by the complete absence of such feature and, consequently, by the complete absence of head-movement in the grammatical system, as notably shown by languages such as Mandarin Chinese. Proceeding further down the word structure hierarchy, the lower parametric options encompass less and less generalized instances of head-movement which in turn correspond to a number of microparametric settings affecting lexically definable subclasses of functional heads like, in case the [+V] option is selected, modal auxiliaries and lexical verbs.

As far as argument drop is concerned, its emergence as a syntactic property is determined by the same hierarchy determining the occurrence of null arguments which has been discussed in section 4.2.1. As Biberauer *et al.* suggest, the grammar of Mohawk selects the second maximally unmarked option even in this context, thus requiring that both functional heads T and v be specified as having an extra unvalued  $\phi$ -feature which, when establishing a defective Agree relation with a matching DP goal, gives rise to what superficially appears to be subject- or object-drop (cf. Biberauer *et al.* 2014, p. 123).

Finally, regarding the last feature of the polysynthetic type, Biberauer *et al.*'s proposal in this respect is that free word order represents the first parametric option which is in turn provided by the parameter hierarchy which determines A'-movement. The formulation proposed here is taken from Roberts (2016) for its greater concision and updatedness (Roberts 2016, p. 187):



According to Biberauer *et al.*, the parametric options specified in (48) concern what is usually referred to as “discourse configurability”, that is, the linguistic property whereby «primary sentence articulation is motivated by discourse-semantic, rather than theta role or case, considerations» (Kiss 1995, p. 3). In this respect, the authors start from the assumption that «phase-heads define local domains, license movement to and/or through their left periphery, and trigger A'-movement» (Biberauer *et al.* 2014, p. 116), the latter being the product of the presence of an optional *edge-feature* (EF), that is, the same feature which, according to Chomsky (2008), allows lexical items to undergo Merge:

For an LI to be able to enter into a computation, merging with some [syntactic object] SO (and automatically satisfying SMT), it must have some property permitting this operation. A property of an [lexical item] LI is called a *feature*, so an LI has a feature that permits it to be merged. Call this the *edge-feature* (EF) of the LI. If an LI lacks EF, it can only be a full expression in itself; an interjection. (Chomsky 2008, p. 139)

Supposing further that «there is universal functional pressure for systems to encode focalization/topicalization» by moving elements into Topic/Focus positions in the left-periphery, cross-linguistic variation in discourse configurability/A'-movement can be accounted for by the effects of the presence of an EF on the phase-heads C, D and v interacting with the locality conditions deriving from Phase Theory (Biberauer *et al.* 2014, p. 116). Assuming Biberauer, Holmberg & Roberts's (2014) proposal that “roll-up” movement of the complement to a position asymmetrically c-commanding the head «is always triggered by a diacritic that they write as “^”» which, if associated with [EF], allows successive-cyclic movement *through* a phase head's edge instead of simply *to* its edge, this theory of discourse configurability yields these four options (the following typology is adapted from Roberts 2016, p. 189):

- (49)
- a. C [EF], v [EF, ^], D [EF, ^] – “free word order”
  - b. C [EF], v [EF, ^], D [??] – scrambling, wh in situ, topicalization
  - c. C [EF], v [^], D [EF] – no scrambling or subextraction
  - d. C [EF], v [EF, ^], D [EF] – like English but with scrambling

Following (49), languages in which all phase-heads host an EF have very liberal scrambling and allow subextraction from nominals, thus enabling constructions in which a modifier occurs disjoint to the noun it modifies (cf. Biberauer *et al.* 2014, pp. 117-118). This

is precisely the case for Mohawk, which therefore represents a maximally unmarked system in this sense, together with the non-polysynthetic Warlpiri. Regarding the remaining options, which are not directly concerned with the immediate thread of this thesis, the next least marked language type corresponds to languages such as Japanese and Korean, which «have quite liberal scrambling, but no clausal-level overt wh-movement in interrogatives» (Biberauer *et al.* 2014, p. 118). Finally, as noted by Roberts, «type (c) includes German and Dutch, while type (d) includes English, North Germanic and the Romance languages» (Roberts 2016, p. 187).

In conclusion, while the “emergent” approach to polysynthesis outlined above is rather tentative if compared, for example, with the approach to pro-drop adopted in Roberts & Holmberg (2010), Biberauer *et al.*’s (2014) proposal represents, in principle, a viable account of the generalized incorporation phenomena occurring in languages such as Mohawk and discussed in Baker (1996).

#### 4.2.5 – *The overt vs covert wh-movement parameter*

Turning now to the issue of cross-linguistic variation in the domain of wh-questions/movement treated by Huang (1982), this classic problem has been debated since the establishment of the P&P framework and is still being debated in today’s parametric theory. As mentioned in the previous paragraph, the A’-movement hierarchy set forth in Biberauer *et al.*’s (2014) represents one plausible possibility to deal with, among other things, the observed duality of overt vs. covert movement. However, although Biberauer *et al.*’s recent proposal fits perfectly well with the emergent parametric model of Roberts & Holmberg (2010) and, hence, with the general account of linguistic variation whereby all the GB parameters which have so far been reviewed in this chapter have been lately reformulated, in this section a different account of variation will be considered, both for the sake of completeness and, ultimately, of some important remarks which will be made in the fifth and final chapter of this thesis. While not completely excluding the idea that variation can also reside in the functional lexicon, such an approach is instead driven by the idea that, as envisioned by Berwick & Chomsky (2011), «parameterization and diversity [...] would be mostly – possibly entirely – restricted to externalization», that is, at the interface between the narrow language faculty and the sensori-motor system (Berwick & Chomsky 2011, p. 37).

Regarding the observed optionality of *wh*-movement, this alternative perspective to linguistic variation is precisely the one adopted in the third chapter of Norvin Richards's work *Uttering Trees* (2010), whose title is *Beyond strenght and weakness*. As the title suggests, Richards's aim is to find an alternative – and actually *explanatory* – account of what has been the mainstream minimalist approach to overt vs covert movement since Chomsky (1992), that is, the postulation of strong or weak features. As noted by the author:

In this type of theory, the overt/covert distinction for *wh*-movement cannot be explained; it is simply stipulated, without following from anything else. Of course, this could turn out to be the right approach. In this chapter I will try, nevertheless, to find a deeper explanation; the goal here will be to predict whether a given language has *wh*-movement or *wh* in situ (or both). (Richards N. 2010, pp. 143-144)

Rather than relying on some kind of feature-specification requiring a given set of *wh*-elements to either appear in a syntactic position adjacent to the complementizer or remain *in situ*, Richards's hypothesis is that «we can predict what a language will do with its *wh*-phrases from the position of its complementizer (particularly, the complementizer associated with *wh*-questions) and the nature of its mapping of syntactic structure onto prosody» (Richards N. 2010, p. 144).

In order to provide the general outline of his own proposal, Richards begins by giving a comparison of the different prosodic profiles of the following Japanese examples, which consist in an affirmative sentence and its corresponding *wh*-interrogative form (*ibidem*):

- (50) a. Naoya-ga    nanika-o            nomiya-de nonda. (Ishihara 2003)  
          Naoya-NOM something-ACC bar-LOC    drank  
          “Naoya drank something at the bar.”  
       b. Naoya-ga    nani-o        nomiya-de nonda no?  
          Naoya-NOM what-ACC bar-LOC    drank Q  
          “What did Naoya drink at the bar?”

More precisely, what Richards is interested in pointing out with respect to the two sentences in (50) is the relative difference in prosodic weight between the direct object *nanika-o* in (50a) and its *wh*-counterpart *nani-o* in (50b) on the one hand, and the general



change affecting the prosodic domain which goes from *nani-o* to the *wh*-complementizer *no* in (50b) on the other. Firstly, compared to the direct object in (50a), according to the author «the *wh*-word direct object in [50b] has its pitch boosted»; secondly, in (50b) «there is a domain, starting with the *wh*-phrase and ending with the *wh*-complementizer [...] which is characterized by pitch compression» but which is absent in (50a) (Richards N. 2010, p. 145). Starting from these observations, which can in turn be read as reflecting the fact that, in Japanese *wh*-questions, a low-level prosodic domain is established between the *wh*-element and the complementizer, Richards's idea is that «every language tries to create a prosodic structure for *wh*-questions in which the *wh*-phrase and the corresponding complementizer are separated by as few prosodic boundaries as possible» (*ibidem*). In Chomsky's (2008) terms, this phenomenon can aptly be thought of as the effect of a PF-related interface condition requiring the *wh*-phrase and the complementizer to be pronounced inside the same prosodic phrase, as shown in (51) (*ibidem*):

$$(51) \quad [\varphi \text{ C } \text{ wh }]$$

In these terms, language variation in the domain of *wh*-questions/movement arises as there is no single method to satisfy the prosodic configuration in (51), but rather two equally viable alternatives which represent, *ceteris paribus*, two equally plausible solutions. Let us suppose to have a prosodic phrasing in which the *wh*-phrase and the complementizer are separated by one or more prosodic boundaries, as in (52) (*ibidem*):

$$(52) \quad \text{C } [\varphi ] [\varphi ] [\varphi \text{ wh}]$$

Given this initial scenario, one way to create a configuration similar to that in (51) is to delete all the prosodic boundaries intervening between the *wh*-phrase and the complementizer – an option which corresponds to covert movement. The other possible way of restructuring (52) is, on the other hand, to overtly move the *wh*-phrase towards the complementizer across all the intervening prosodic boundaries, thus creating a prosodic configuration analogous to (53) (*ibidem*):

$$(53) \quad [\text{wh C } [\varphi ] [\varphi ] [\varphi \text{ wh}]]$$

According to Richards, whether a language selects overt or covert *wh*-movement can be predicted by taking into account two independent properties that characterize the

prosody of the language in question, namely, the placement of Minor Phrase boundaries and the position of complementizers with respect to their complements. Starting from the assumption that «prosodic representations are constructed by mapping certain syntactic boundaries onto prosodic boundaries», the first property mentioned by the author concerns which of the two syntactic edges of every DP is associated with a boundary in the prosodic domain (Richards N. 2010, p. 148). For example, the following representation shows how the prosodic structures of a typical SOV sentence are derived by selecting the left edge of every DP, being it either wh- or non-wh (Richards N. 2010, p. 150):

- (54) a. [TP [DP whP ] [VP [DP D NP ] V ]]  
 b. ( whP )( D NP V )

While (54a) corresponds to the sentence's syntactic representation, (54b) corresponds to what the literature on prosody refers to as the *Minor Phrase*, that is, «the lowest level of phonological phrasing» (Richards N. 2010, p. 148). As noted by the author, one important property of Minor Phrases is that they may combine recursively to form larger Minor Phrases (sometimes called Major Phrases) according to the following algorithm (Richards N. 2010, p. 150):

- (55) a. For one end of the larger Minor Phrase, use a Minor Phrase boundary that was introduced by a *wh*-phrase.  
 b. For the other end of the larger Minor Phrase, use any existing Minor Phrase boundary.

Considering the example given in (54), by applying the algorithm in (55) to the phrasing in (54b) we obtain a larger Minor Phrase consisting, on the one hand, of the Minor Phrase boundary that was introduced by the left edge of the wh-phrase and, on the other hand, of the right boundary at the end of the utterance, thus yielding the higher-level prosodic phrasing in (56c) (*ibidem*):

- (56) a. [TP [DP whP ] [VP [DP D NP ] V ]]  
 b. ( whP )( D NP V )  
 c. ( whP D NP V )

Crucially, the point of this procedure is that, according to Richards, the algorithm which has been applied to (56a) in order to derive (56c) is exactly the means by which

every *wh-in situ* language is able to create a “wh-domain” encompassing the wh-phrase and the complementizer, thus allowing them to share a Minor Phrase and to meet the following condition on wh-prosody (Richards N. 2010, p. 151):

- (57) Given a *wh*-phrase  $\alpha$  and a complementizer C where  $\alpha$  takes scope,  $\alpha$  and C must be separated by as few Minor Phrase boundaries as possible, for some level of Minor Phrasing.

Conversely, all those languages which do not distribute their Minor Phrase boundaries in such a way that there still is one or more prosodic boundaries between the wh-phrase and the complementizer will feature *wh-ex situ*, here meant as a strategy for overtly realizing these elements in adjacent positions. Crucially, the choice each language makes about which edge of its DPs is being mapped onto a Minor Phrase boundary interacts with the position of complementizers, thus yielding four parametric combinations:

There will be two main points of crosslinguistic variation that will be relevant for us. One will be the position of the complementizer: complementizers may either precede or follow their complements. The other has to do with the placement of Minor Phrase boundaries, which can be either to the Left or to the Right of certain maximal projections. These two binary parameters leave us with four logical possibilities, which we will spend the rest of this section outlining schematically. (*ibidem*)

The first combination is the one represented by «a language with a final complementizer that places Minor Phrase boundaries at Left edges of certain maximal projections» (Richards N. 2010, pp. 151-152). As shown above, Japanese corresponds to such a description. Taking into consideration the position of the wh-complementizer *no*, the application of the algorithm in (55) to sentence (50b) yields the wh-domain shown in the following schema (Richards N. 2010, p. 152):

- (58) a. [<sub>DP</sub>] [**whP**] [<sub>DP</sub>] V C  
       b. (    ) (    ) (    )  
       c. (    ) (    ) (    )

As such language type is capable of parsing the wh-phrase and the complementizer in a single prosodic domain, it is predicted to have *wh-in situ*. This is indeed the case for Japanese.

The second language type corresponds to a language which is complementizer-final but which «marks not Left edges, but Right edges of certain maximal projections with a Minor Phrase boundary» (Richards N. 2010, p. 152):

- (59) a. [DP] [**whP**] [DP] V C  
 b. ( ) ( ) ( ) ( )  
 c. ( ) ( ) ( )

Given the above premises, with the complementizer occurring in clause-final position, we would have to conclude that such a language would be unable to have *wh-in situ*, and this because the algorithm in (55) could never be able to eliminate the right prosodic boundaries occurring between the *wh*-phrase and the complementizer. As noted by the author, this is precisely what happens in Basque, a language which obligatorily moves the *wh*-phrase in immediate preverbal position, as shown by the grammaticality contrast between (60a-61a) and (60b-61b) (Richards N. 2010, p. 162):

- (60) a. Mirenek séin ikusi rau? (Arregi 2002)  
 Miren-ERG who-ABS see-PRF AUX.PR  
 “Who has Miren seen?”  
 b. \*Séin Mirenek ikusi rau?  
 who-ABS Miren-ERG see-PRF AUX.PR  
 (61) a. Jon señek ikusi rau?  
 Jon-ABS who-ERG see-PRF AUX.PR  
 “Who saw Jon?”  
 b. \*Señek Jon ikusi rau?  
 who-ERG Jon-ABS see-PRF AUX.PR

As far as the third language type is concerned, such systems are those in which «both the complementizer and Minor Phrase boundaries precede *wh*-phrases» (Richards N. 2010, p. 153). Being the mirror image of Basque with respect to these properties, these languages will analogously be forced to resort to *wh*-movement by their inability to improve the prosodic status of *wh*-questions by creating a larger Minor Phrase boundary. The outcome of Richards’ proposed algorithm would in fact be the following (*ibidem*):

- (62) a. C [DP ] [whP ] [DP ]  
 b. ( ) ( ) ( ) ( )  
 c. ( ) ( ) ( ) ( )

As Richards notes, Tatalog is a language with such properties and, strikingly, its wh-questions obligatorily display wh-movement to the left periphery (Richards N. 2010, p. 181):

- (63) a. **Kailan** umuwi si Juan?  
 when NOM-went.home ANG Juan  
 “When did Juan go home?”  
 b. \*Umuwi si Juan **kailan**?

Finally, the fourth and last case of interest corresponds to languages in which «the complementizer is initial, and Minor Phrase boundaries follow their maximal projections» (Richards N. 2010, p. 153). Analogously to Japanese, of which they represent the mirror image, these languages ought to have what it takes to be able to neutralize the prosodic boundaries intervening between the wh-phrase and the complementizer and, consequently, to restructure the latter two elements into one single prosodic unit, as shown in the following representation (Richards N. 2010, p. 154):

- (64) a. C [DP ] [whP ] [DP ]  
 b. ( ) ( ) ( ) ( )  
 c. ( ) ( ) ( ) ( )

One language which has sentence-initial complements and marks right edges of certain XPs with prosodic boundaries is Chichewa. According to the hypothesis defended by the author, this language should allow wh-*in situ* in wh-questions. This speculation is correct, as showed by the following example (Richards N. 2010, p. 184):

- (65) anaményá chiyáani ndi mwáálá. (Downing 2005)  
 he.hit what with rock  
 “What did he hit with the rock?”

Despite of the fact that this collection of examples is not enough to demonstrate the existence of a universal PF well-formedness condition on *wh*-questions, especially considering that, as Richards himself admits, «the theory will clearly have to be tested against a wider range of languages», what is actually important here is the general idea behind this proposal, namely, that there can be variation at the interface between narrow syntax and the articulatory-perceptive system (Richards N. 2010, p. 187). Crucially, this assumption implies that the traditional idea that all parametric variation is confined to the lexicon in the form of different feature specifications on functional categories is either wrong or that, assuming we do not want to eliminate syntactic parameters but rather go beyond them, some instances of variation are not actually parametric in nature, but reside outside the lexicon. In particular, it is worth mentioning that in Richards N. (2010) *wh*-movement is not referred to as a parameter but rather as an *apparent parameter* deriving, in turn, from other parameters. This can be seen in the following quotation:

Ultimately, the hope is to apply this way of thinking to other types of movement, as well. We have grown accustomed to being able to stipulate that this or that type of movement (not only *wh*-movement, but also scrambling, head movement of the verb to T, and so on) is present or absent in a given language. The idea here has been to derive this apparent parameter from other parameters, just in the case of *wh*-movement. (Richards N. 2010, p. 200)

Although this characterization of *wh*-movement could well be true, this is just speculation since a lexically-based account in the spirit of Biberauer *et al.* (2014) could actually turn out to be the right one; yet still, regardless of this particular case, given a scenario in which a language is put in front of two alternatives and, crucially, is not given any bias towards (or against) any one of them by neither UG or third factor considerations, whenever the final decision rests with a PF-interface condition (and not with a binary feature driving the narrow-syntactic derivation) we would indeed find ourselves in front of a kind of optionality which, although systematic, could not be considered parametric on a par with all the other instances of variation which conform to the “canonical” notion of parameter put forth in Borer (1983).

## 4.2 – Conclusion

This chapter has aimed at evaluating the main parameters of the GB Theory in the light of contemporary generative grammar by putting them into a minimalist theoretical context. According to their compatibility with current Generative Grammar, it has been shown that a good half of them have now been discarded (either for their manifest inadequacies or, simply, as their primary function has been replaced by conceptually “lighter” devices), with the remaining ones being now accounted for almost exclusively by the lexically-based, hierarchical parametric model elaborated by Roberts & Holmberg (2010) and further developed in Biberauer *et al.* (2014). That being said, although in general the parameters of GB Theory which have stood the test of time have undergone some major changes with the advent of the MP and the consequent necessity of a minimally specified language faculty in the narrow sense, what emerges as most significant from the evaluation carried on in this chapter is that, strikingly, the only traditional parameters which still enjoy an independent theoretical status are those which have been classified as Spellout parameters. Moreover, as hinted at above, among this parametric class there could well be an exception in this sense, although initially characterized as a parameter belonging to this same class: the overt vs covert movement parameter. In fact, assuming Richards N. (2010) or an equivalent PF-based account to be on the right track, the observed duality of overt vs. covert wh-movement would no more be conceivable as a syntactic parameter, but rather as a choice being made right before Spellout in order to satisfy a specific constraint imposed at the interface between syntax and the articulatory-acoustic system.

Leaving aside for our present purposes the V-to-C movement parameter, which has been argued to correspond to an idiosyncratic, non-parametric property of the same kind as S'-deletion, delving further into the possibility that there could actually be a principled difference between the overt vs covert wh-movement parameter and the remaining Spellout parameters, some distinctions ought surely to be made. Most notably, if on the one hand it is true that the incorporation-driven parameters epitomized by the null subject parameter as reformulated by Roberts (2010) «are also plausibly derived from more primitive conditions on spell-out and linearisation» (Holmberg 2010a, p. 98, n. 7), on the other hand the mechanism behind Richards N.'s (2010) approach is radically different. In the first case, head-movement is triggered by the *optional* presence of an unvalued morphosyntactic feature which, when specified on a given probe, determines an Agree

relation between a probe and a defective goal, with the result that it is *invariably* the lower copy that is taken care of by the PF-deletion process of chain reduction. In the second case, however, wh-movement is assumed to be an inherent property of all human languages: the only change happening regards, crucially, which one of the who copies are pronounced, with this choice being made outside narrow syntax by a PF-constraint. Although this kind of optionality at the syntax-phonetics interface does not overlap with the P&P model, which rather assumes variation to take place solely in the lexicon, it is still congenial to a minimalist view of language faculty which has to be optimally designed in order to meet the requirements of «a sensorimotor system which is there, independently of the language; maybe [...] somewhat modified because of the presence of language, but in essence it is there independently of language» (Chomsky 2002, p. 108). As noted by Berwick & Chomsky (2011):

Externalization is not a simple task. It has to relate two quite distinct systems: one is a sensori-motor system that appears to have been basically intact for hundreds of thousands of years; the second is a newly emerged computational system for thought, which is perfect insofar as the strong minimalist thesis is correct. (Berwick & Chomsky 2011, p. 37)

In these terms, if on the one hand «the essential property of language must be that it satisfies the interface conditions» in order to be usable at all, on certain occasions it could be conceivable that language has no single solution to do so, or rather, that narrow syntax does not make available one single representation which perfectly meets the legibility conditions imposed by the sensorimotor system (Chomsky 2002, p. 158):

Suppose that a super-engineer were given design specifications for language: “Here are the conditions that FL must satisfy; your task is to design a device that satisfies these conditions in some optimal manner (the solution might not be unique).” The question is, how close does language come to such optimal design? (Chomsky 2000, p. 92)

Therefore, variation at the PF-branch of grammar could hence be a consequence of the fact that, out of a limited set of equally less-than-perfect possibilities, each language chooses the most optimal one in terms of legibility conditions according to its own interface-related properties. Such an eventuality arguably has to be considered, as not only can it not be excluded *a priori*, but it also is strongly desirable in light of a view of language faculty which is not designed to meet the specific conditions – that is, the physical limits – imposed by the external systems.



## Chapter V

### The head-complement parameter

#### 5.1 – The first explicit parameters of word order

It is arguably not by chance that perhaps the best known parameter of all is the head-complement parameter. Although it was only from the early Eighties onwards that, on the wave of Rizzi and Taraldsen's pre-parametric inquiries, the whole generative enterprise become interested in language typology, the availability of extensive cross-linguistic data which had been provided by previous typological studies – most notably by Greenberg's (1963) seminal work on language typology – soon drew Generative Grammar's increasing attention to a number of word order correlations which immediately seemed too striking to be merely accidental. In particular, the fact that Greenberg's so-called "harmonic orders" constituted visible manifestations of the existence of implicational relations between cross-categorical patterns of directionality seemed to hint at the existence of that rich deductive structure at the basis of language acquisition and cross-linguistic variation which would soon be argued for in Chomsky (1981a).

##### *5.1.1 – The parametrization of recursion directionality*

A perfect example of Generative Grammar's urge to account for «the order of head with respect to complement across the different constituents – while it varies cross-linguistically – is (approximately) identical within one and the same language» is the paper '*Universalì di Greenberg' e grammatica generativa* ("Greenberg's universals' and generative grammar") by Graffi (1980) (Graffi 2001, p. 459). As the title suggests, the primary aim of this work was to propose an account of the generalizations regarding the relative ordering of major sentential elements (Subject, Verb and Object) with respect to the head-complement order across phrasal categories which had been proposed by Greenberg (1963) in a way which could be compatible with both the X-bar theory of phrase structure of Chomsky (1970) and the P&P approach to linguistic variation outlined in Chomsky (1981a).

In this paper, Graffi introduces his proposal by focusing first of all on the main claim of X-bar theory, namely, that words associated with the four main lexical categories N, A,

V and P always project the same syntactic structure, which in turn can be represented by means of the following rewriting rules:

(1)  $X'' \rightarrow \text{Spec } X'$

(2)  $X' \rightarrow X \text{ Compl}$

The central idea put forth by Graffi is that, if on the one hand (1) and (2) show that the relation between a head and its specifier and that between a head and its complement can not be parameterized without invalidating the very core of X-bar theory, on the other hand «what has to be parameterized [...] concerns the order of the Complement and the Specifier with respect to the head» (Graffi 1980, p. 374; my translation). If the opposite were true, with phrase structure rules (1) and (2) determining, in addition to immediate dominance relations within the phrase, the fact that the head invariably precedes its complement and follows its specifier, it would in fact be impossible to have languages which have OV as their regular word order, as «it is evident, in fact, that in OV languages the Complement follows the head rather than precedes it» (Graffi 1980, p. 375; my translation). As noted by Greenberg (1963), the basic order OV is a staple of Turkish, a language which displays the head-complement pattern not only in VPs but, strikingly, also in APs, NPs and genitive PPs:

Linguists are in general familiar with the notion that certain languages tend consistently to put modifying or limiting elements before those modified or limited, while others just as consistently do the opposite. For an example of the former type, Turkish puts adjectives before the nouns they modify, places the object of the verb before the verb, the dependent genitive before the governing noun, adverbs before adjectives which they modify, etc. Such languages, moreover, tend to have postpositions for concepts expressed by prepositions in English. (Greenberg 1963, p. 60)

Looking for a feasible account the systematic cross-linguistic patterns of word order variation mentioned above, a first possibility in this respect could be to assume, following Van Riemsdijk (1978), that «there may be a universal tendency to have the complements and the specifiers on the opposite sides of the head, a tendency that emerges quite clearly from Greenberg's (1966 [=1963, A.R.]) observations» (Van Riemsdijk 1978, p. 124, n. 2). However, as noted by Graffi, while Van Riemsdijk proposal allows, interestingly, a simple binary parametrization of X-bar theory by setting either the sequence Specifier-Head-Complement or the sequence Complement-Head-Specifier as the possible linear orders

across the different constituents of a language, once specifiers are taken into account the picture changes drastically. Although it is true that Greenberg (1963) did not formulate an explicit distinction between specifiers and complements, the application of X-bar theory to a relevant set of typological data shows that, as pointed out by Graffi with respect to the relative order between the noun and its modifying demonstrative on the one hand and between the noun and its modifying numeral on the other, «all “rigid” OV languages (namely, those which always have the verb in clause-final position: Burmese, Burushaski, Kannada, Japanese and Turkish, in Greenberg’s sample) put all Specifiers on the same side of Complements, that is, to the left of the head» (Graffi 1980, p. 376; my translation).

Given the descriptive inadequacy of Van Riemsdijk’s attempt to “parametrize” the linear order of the specifier and the complement with respect to the head of the phrase, Graffi proposes to capture «the undeniable cross-linguistic tendency to place Complements on the same side of the head in all four major phrases» and «the apparently more bizarre behaviour of Specifiers» displayed in languages such as Turkish by means of two distinct parameters (Graffi 1980, p. 377; my translation). According to the author, however, the aspect of UG which allows parametrization does not lie in X-bar theory itself but, from a broader perspective, in how major phrases realize what Chomsky (1965) calls «the true recursive property» of the base component, that is, the possibility of any natural language to embed an indefinite number of propositions inside another (Chomsky 1965, p. 225, n. 11):

Now the recursive property is a feature of the base component, in particular, of the rules that introduce the initial symbol S in designated positions in strings of category symbols. There are, apparently, no other recursive rules in the base. (Chomsky 1965, p. 137)

In these terms, the fact that all major phrases (VP, NP, AP and PP) can have an S-node as their complement but not as their specifier, and hence that «the side of the unmarked appearance of complements coincides with that of embedded clauses with respect to the main one» suggests the existence, according to Graffi, of a parameter which sets the recursive side selected by each language (Graffi 1980, p. 379; my translation):

(3) Every language fixes the recursive side of all major phrases with respect to the head.

At first glance, the parameter formulated in (3) would seem insufficient to explain the inconsistencies affecting those languages which, for example, although being

classified as VO, also exhibit some patterns generally associated with the OV type. This had in turn been exemplified in Greenberg (1963) by the behaviour of English which, although being a VO language, features the orders Adjective-Noun and Genitive-Noun:

The majority of languages, as for example English, are not as well marked in this respect. In English [...] there are prepositions, and the noun object follows the verb. On the other hand, English resembles Turkish in that the adjective precedes the noun. Moreover, in the genitive construction both orders exist: "John's house" and "the house of John". (Greenberg 1963, p. 60)

Given these premises, the solution proposed by Graffi is to rely on the notion of markedness. Far from being an *ad-hoc* device to save the parameter, markedness is here conceived as a crucial property of language which, as such, may be called upon only provided there is some principled reason to do so. In the present case, the criterion for markedness adopted by Graffi consists not only in the fact that the existence of a marked option in a given grammatical system necessarily implies the existence, in the same system, of the corresponding unmarked option, but also that the recursive property referred to in (3) only pertains to those constructions belonging to the unmarked type. As far as English is concerned, the fact that the Adjective-Noun order represents the marked counterpart of the Noun-Adjective order is showed by sentences such as (4) (Graffi 1980, p. 380):

(4) I climbed [<sub>NP</sub> a mountain [<sub>AP</sub> higher [<sub>PP</sub> than I had ever climbed before ]]].

Although the first order is more frequent than the second one, only in those cases in which «the Adjectival Phrase is formed not by the head alone» – that is, not in those cases in which the adjective immediately precedes the noun it modifies – «but by the head plus the Complement» is the adjectival head able to introduce an embedded clause on its right (*ibidem*; my translation). By the same reasoning, according to Graffi, «marked cases are also those corresponding to pronominal genitives (the so-called "saxon genitives")», and this because only post-nominal genitives can realize the recursive property of the language in the sense of Chomsky (1965) (*ibidem*; my translation).

Turning now to what invalidated Van Riemsdijk's conjecture, that is, the placement of specifiers with respect to the complements, Graffi proposes this second parameter (Graffi 1980, p. 382; my translation):

- (5) Some languages use exclusively the recursive side with respect to the head; the other ones use also the opposite side.

As a consequence of (3) and (5), only specifiers are expected to occupy the non-recursive side of the head, while, conversely, complements are expected to occupy only the recursive side. Similarly to what argued in the discussion of the “recursive side parameter”, also this second parameter admits certain marked violations. In this sense, the violations assumed by the author can arise on two specific cases: first, when some specifiers happen to occupy the non-recursive side even if this side is not actually specified as available; and second, concerning those languages which allow both sides of their heads to be used, when there are specifiers on the recursive side (cf. Graffi 1980, p. 382). While not delving into the details of these matters, according to Graffi such violations can nonetheless somehow be accounted for from a markedness perspective.

Summing up the previous arguments, what Graffi’s proposal aims to convey is that the essential difference between complements and specifiers within a generative view of language lies not only in their respective structural positions within the phrase but also, at an even deeper level, in their distinct role with respect to recursion, which is here assumed to be a defining property of complements since only these latter categories can host a subordinate clause. Consequently, the possible variability in the linearization patterns followed by these two syntactic categories has not to be regarded as an independent property, but rather as a consequence of the fact that language faculty incorporates a rich internal deductive structure which allows systematic variation in a way that can be assimilated to markedness effects. This aspect explicitly emerges in the author’s replies to a hypothetical objection concerning the impossibility for grammatical theory, here referred to with the same systematic ambiguity as in Chomsky (1965), to attribute the collocation of a specifier on the non-recursive side to either a marked violation of (5) or the direct effect of this parameter’s positive setting. As the author notes:

In the first place, it should be borne in mind how Greenberg’s data about Specifiers (although largely incomplete, as exclusively limited to the Specifiers of N”) show clearly enough the tendency of some languages to use only the recursive side and, at the same time, of some others to use both sides. The observed regularity, in either one sense or the other, would then lead us to consider these phenomena as the effects of a parametric choice. (Graffi 1980, p. 383, my translation)

From another perspective, which nonetheless does not conflict with, but rather, complements this first answer, Graffi points out how a sufficiently articulated theory of grammar should be able to distinguish, between maximally unmarked syntactic configurations and maximally marked ones, a set of different options characterized by an intermediate degree of markedness. For example, languages violating the first parameter in a consistent way – that is, without admitting any complement at their selected recursive side – would still represent marked systems but, differently from all of those languages admitting less harmonic linearization patterns, would represent minimally marked systems with respect to (3):

The second answer that can be given to our objection is that the interpretation of a grammar as a “highly differentiated system” can go well beyond the simple dichotomy ranging from “central phenomena” on the one hand and “peripheral” or “marked” ones on the other: between the centre and the periphery one can postulate the existence of large intermediate areas. The phenomena we are dealing with could indeed be located in one of such areas: languages using the non-recursive side in a systematic way could represent an instance of marked phenomenon, but, as to say, “on a low degree of markedness”. (Graffi 1980, p. 383, my translation)

In conclusion, although this work does not pretend to be a treatise on the subject, Graffi’s (1980) proposal of a parameterization of recursion directionality certainly represents, given its background and the time of writing, an ingenious way to account for the cross-categorical harmony in directionality within each language and various potential counterexamples which had been identified throughout the data provided in Greenberg (1963).

### *5.1.2 – The elimination of the categorial component of the base and the postulation of the head-complement parameter*

Although a historical review of a theoretical proposal such as that concerning the existence of a parameter accounting for cross-linguistic variation in word order phenomena must necessarily retrace the most important formulations of the said parameter, another aspect which should also be referred to in such enterprise consists in those theoretical advancements which paved the way for subsequent parametric formulations. In this respect, a crucial step in the development of the head-complement parameter and, more

in general, of Generative Grammar as a theory of language itself has been represented by Stowell's PhD dissertation *Origins of Phrase Structure* (1981).

One, if not the main, goal pursued in Stowell's thesis was, in the spirit of the new theoretical approach which had been carried out from Chomsky (1981a) onward, to get rid of all those limits and redundancies which could compromise Generative Grammar's capability of providing complete explanatory adequacy for linguistic phenomena. Within the tradition outlined in *Aspects of the Theory of Syntax* (Chomsky 1965), these two factors were both represented by the categorial component of the base. According to the original model of Generative Grammar, the phase which had been referred to as the *standard theory*, this system consisted of a set of language-specific, context-free rewrite rules whose expansion encoded both hierarchical structure and linear order. For instance, a base rule for a verb phrase as in example (6) states that a verb phrase is syntactically expanded to form a string consisting of a verbal head and an optional NP (cf. Travis 1984, p. 10):

$$(6) VP \rightarrow V (NP)$$

Given the fact that these rules define in advance the internal structure of the term appearing to the left of the expansion, the existence of such powerful devices would allow language faculty to encode too detailed – and, worst of all, idiosyncratic – syntactic information. In these terms, as stated by Stowell, «the theory of syntax has very little to offer to the theory of acquisition in the way of a set of predetermined hypotheses about what kind of detailed information to expect», as the descriptive adequacy attainable by such rules would largely overcome explanatory adequacy (Stowell 1981, p. 71).

In addition to the considerations outlined above, the categorial component represents an issue to deal with also with respect to the format itself of Chomsky's linguistic theory. In *Remarks on Nominalization* (Chomsky 1970), Chomsky proposed that lexical heads carry a piece of information known as subcategorization frame – namely, a set of categorial features which specify the number and type of the syntactic arguments with which the lexical item needs to co-occur (cf. Chomsky 1993 [1981a], p. 35). It goes without saying that, under such assumptions, the existence of categorial rules turns out to be an unbearable redundancy within the overall structure of the grammar, as the same information these rules would specify about any phrase would also be encoded in lexical heads of the latter (cf. Stowell 1981, p. 71). Therefore, as well expressed by Stowell:

[...] the theory of phrase structure implicit in this rule system is largely redundant, offers no real depth of explanation, and provides little more than an arbitrary collection of observed generalizations about each phrasal category. (Stowell 1981, pp. 50-51)

Thus, given all these methodological problems, Stowell suggests that «the categorial component does not exist» (Stowell 1981, p. 51). Building on Hale's (1980) proposal that, in non-configurational languages, all base rules are essentially formulated in category-neutral terms, thus meaning that «all rules of phrase structure consistently generalize over all parts of speech», Stowell extends this hypothesis also to configurational languages such as English (Hale 1980). What this means is that, with phrase structure rules being therefore unable to refer to categorial features, what we are left with corresponds to the essential properties of X-bar theory, here conceived as the maximally general set of constraints on phrase structure rules given below (Stowell 1981, p. 70):

- (7) a. Every phrase is endocentric.
- b. Specifiers appear at the X'' level; subcategorized complements appear within X'.
- c. The head always appears adjacent to one boundary of X'.
- d. The head term is one bar-level lower than the immediately dominating phrasal node.
- e. Only maximal projections may appear as non-head terms within a phrase.

As pointed out by the author, this conception of X-bar theory is fully compatible with the approach to language acquisition put forth in Chomsky (1981a) and involving «a theory of a core grammatical structure, most of which holds constant across all languages, except for parameters with easily-identified empirical effects [...] that are left open at various points in the deductive structure of the grammar» (Stowell 1981, p. 73). Since the characterization of X-bar theory given in (7) identifies specific points of predetermined structure which do not encompass the linear position of the head term at the intermediate phrase level X', Stowell suggests that the distinction between SOV languages and SVO languages can be derived by parameterizing the position of the head with respect to its subcategorized complements according to the possible settings shown in (8) (Stowell 1981, p. 74):

- (8) a.  $V' \rightarrow \dots V$
- b.  $V' \rightarrow V \dots$



In these terms, therefore, the possibility of accounting for cross-linguistic variation in the linear position of the head with respect to its complements would be a further desirable consequence of the elimination of the categorial component.

If on the one hand it has been shown that there are strong theoretical reasons to dispose of the categorial component altogether, on the other hand the elimination of phrase structure rules indeed seems to represent a step backwards in terms of descriptive adequacy, as «this would make it impossible for the rule system of the base to stipulate differences among the various categories with respect to external distribution of internal structure» (Stowell 1981, p. 85). This will be especially true for those languages which, like English, if on the one hand display a rather fixed word order, on the other hand show some cross-categorial variation in the internal structure of their major phrases:

Unfortunately, the theory of phrase structure makes no real predictions beyond the realm of X-bar theory. In particular, the categorial identity and mutual ordering of complements and specifiers within each level is left completely open to random cross-linguistic variation. Virtually any string of phrases can appear in the expansion of any given rule, so there is simply not enough predetermined organization of terms for specific aspects of the rule expansions to be identified as isolated points in the structure that serve as the variables left open for parametric variation. (Stowell 1981, p. 75)

However, what Stowell argues at this point is that descriptive adequacy is, in spite of these premises, not at stake, and that the fact that «the elimination of the categorial component forces very specific analyses of a number of constructions» has to be regarded not as an undesirable shortcoming but rather as a clear advantage for Generative Grammar's potential for achieving a maximum degree of descriptive adequacy without generating a tension with the conditions of explanatory adequacy (Stowell 1981, p. 86). In this terms, according to the author:

[...] the constellation of phenomena traditionally associated with language-specific phrase structure rules can be deduced from the interaction of general principles of the language faculty with specific options left open for parametric variation at certain points in the structure of the grammar. (Stowell 1981, p. 51)

As far as these general principles of the language faculty are concerned, in the third chapter of Stowell (1981) emphasis is given to Case theory and Theta-theory, as these are the main theoretical devices whereby the author attempts to account for the order of

constituents within the level of complement structure in English. As an example of the abandonment of language-specific phrase rules in favour of the more explanatory approach argued for above, the specific aspect briefly discussed here concerns the position of direct object NPs with respect to a head V or P. In this regard, Stowell's argumentation begins by outlining one systematic property of English, namely, that «the object of a verb or preposition must immediately follow its head», as shown in (9-10) (Stowell 1981, pp. 106, 108):

- (9) a. Neil's donating [ten dollars] to the fund was a nice gesture.  
       b. \*Neil's donating to the fund [ten dollars] was a nice gesture.
- (10) a. Ted talked to [his kids] about the war.  
       b. \*Ted talked to about the war [his kids].

Starting from this observation, Stowell first considers the possibility that this restriction can be captured by a specific category-neutral phrase structure rule applying to the phrase consisting only of the head and its closest argument  $\alpha$  (Stowell 1981, p. 108):

- (11)  $X' \rightarrow X - \alpha$  ( $\alpha$  = the closest argument)

Although the above rule easily captures the facts observed in (9-10), there are some cases in which the order of complements is not the one predicted by (11). This is shown, for instance, by the linear distribution of derived nominals and their direct object PPs, as shown in (12) (Stowell 1981, p. 109):

- (12) a. Neil's donation [of ten dollars] to the fund was a nice gesture.  
       b. Neil's donation to the fund [of ten dollars] was a nice gesture.

According to the author, the fact that derived nominals such as *donation* can strand their object is due to the fact that, differently from their verbal counterpart, they do not assign Case. This can be inferred by one crucial difference between derived nominals and gerunds, namely, that «the objects of derived nominals are subject to the rule of *of*-Insertion, whereas verbs take bare NP objects», as shown by (12a-9a), (here repeated as (13a-13b)) (Stowell 1981, p. 110):

- (13) a. Neil's donation [of ten dollars] to the fund was a nice gesture.  
 b. Neil's donating [ten dollars] to the fund was a nice gesture.

Starting from the assumption that, according to (a simplified version of) Chomsky's (1981a) Visibility Condition, « $\theta$ -roles may only be assigned to A-positions which are associated with PRO or Case» (Stowell 1981, p. 111), Stowell hypothesizes that in (14a) the preposition *of* is inserted before the direct object NP because of its function as a dummy Case marker, which is, in turn, the determining factor in allowing for theta-role assignment by the verb to the object NP, in conformity with the Chomsky's (1981a) Theta-Criterion:

[...] derived nominals lack the Case-assigning feature [-N]; therefore they can only assign a  $\theta$ -role to a noun phrase that is assigned Case by some other means; this is why *of*-Insertion is required. Since *of* is a preposition, it bears the feature [-N], and can function as a "dummy" Case-marker. This allows the NP to which it is adjoined to satisfy [the Visibility Condition], making  $\theta$ -role assignment possible, as required by [the Theta-Criterion]. (Stowell 1981, p. 112)

Returning to the required adjacency between the NP object and its governing head, Stowell's proposal is that, in languages like English, the procedure of Case assignment is specified with an additional clause (14ii) requiring linear adjacency between the Case-marking head its governed element:

- (14) In the configuration [ $\alpha$   $\beta$ ...] or [... $\beta$   $\alpha$ ],  $\alpha$  Case-marks  $\beta$ , where  
 (i)  $\alpha$  governs  $\beta$  and  
 (ii)  $\alpha$  is adjacent to  $\beta$ , and  
 (iii)  $\alpha$  is [-N]  
 (Stowell 1981, p. 113)

In these terms, the fact that the direct object NP of a verb or preposition must immediately follow its head follows from the interaction of the Theta-Criterion and the Case Filter (cf. *ibidem*). This also explains why the objects of derived nominals do not need to appear immediately after the latter, as shown in (15) (Stowell 1981, p. 110):

- (15) The notoriety resulting from [Kathy's exposure in the Washington Post [of Nixon's war crimes ]] led to her new assignment.

In these cases, in fact, it is the preposition *of* which, by acting as a dummy Case-marker, satisfies (15ii) and hence allows the head noun to theta-mark its *of*-NP object.

Although Stowell's analysis of the order of complements within the X' level of phrase structure goes on, its further development is beyond the scope of the present thesis. However, what emerges from the formulations seen so far is that, according to his proposal, not only could much of the task traditionally assigned to language-specific phrase structure rules be undertaken by the various subsystems of UG – with Case theory and theta-theory capturing the lion's share in this respect – but also that cross-linguistic variation in the linearization of head-complement structures could be attributed to the setting of a specific binary parameter, thus allowing Generative Grammar to do without a set of highly descriptive explanatory language-specific rules which undermined its explanatory power.

### *5.1.3 – A parameterization of head-directionality, theta-role assignment, Case-assignment, and predication*

If on the one hand the approach followed in Stowell (1981) was not mainly concerned with cross-linguistic variation, the cross-linguistic issue would instead play a most prominent role in the analysis pursued some years later in the doctoral dissertation of Lisa deMena Travis, whose title is *Parameters and Effects of Word Order Variation* (1984). In this work, which ideally represents a continuation of Stowell's effort to demonstrate the expendability of the component of categorial rules which had been inherited from the previous theoretical phase of Generative Grammar, Travis proposes to account for the issue of cross-linguistic variation on linear order by adding three sub-parameters to the one determining head-directionality, namely, the parameters of direction of theta-role assignment, Case-assignment, and predication.

Travis's cross-linguistic inquiry is presented in the second chapter of this work, which mainly consists in a cross-linguistic comparison between Archaic Chinese (AC) and two diachronic stages of Modern Mandarin which are respectively referred to as MM1 and MM2. Starting with AC, the first aspect observed by the author is that, in this language, «objects and prepositional phrases always appear after the verb», as shown in (16) (Travis 1984, p. 40):

- (16) qu yue yu ci. (Light 1979)  
 play music prep here  
 [ V O PP ]

Going from VPs to NPs, the latter constituents in AC are all head-final, as shown by the nontrivial fact that «relative clauses, modifiers, and genitive NPs all precede the head noun» (Travis 1984, p. 40). Finally, PPs are head-initial, as shown in (17) (Travis 1984, p. 41):

- (17) chu yu you gu. (Li & Thompson 1973)  
 emerge from dark valley  
 [ V [ P NP ] ]

In parametric terms, therefore, AC can be described as a head-initial language with respect to VPs, but with head-final NPs since, according to Travis's theory of linearization, «this parameter may vary for different categories» (Travis 1984, p. 90).

Now is the turn of Modern Mandarin (MM1). As noted by the author, «it is in the VPs [...] that one can see the greatest change in word order» (Travis 1984, p. 41); in fact, most PPs now appear preverbally, as shown by the following examples (*ibidem*):

- (18) zai zher yanzou yinyue. (Light 1979)  
 prep here play music  
 (19) cong you gu chulai. (Li & Thompson 1973)  
 from dark valley emerge  
 [ PP V ]

Although MM1 seems to basically conform to the order S-PP-V-O, some further distinctions have to be made. The first one concerns the difference between postverbal PPs and preverbal PPs. In MM1, PPs introduced by the prepositions *gei* “to/for” and *zai* “at” may appear post-verbally as well as preverbally. However, what is particularly interesting about these elements is that their meaning changes depending on their position with respect to the verb. More precisely, as noted by Travis, «*gei* NP before a verb is benefactive while *gei* NP after the verb is dative» (Travis 1984, p. 46):

- (20) ta gei wo mai le chezi le. (Li & Thompson 1975)  
 he for me sell asp car ASP  
 "He sold a car for me."
- (21) ta mai gei wo chezi le.  
 he sell to me car ASP  
 "He sold a car to me."

The same is true for *zai*. With motion verbs, for example, «postverbal *zai* is directional while preverbal *zai* is locational» (Travis 1984, p. 46):

- (22) Zhang-san tiao zai zhuozi-shang. (Li & Thompson 1975)  
 Z. jump at table-on  
 "Zhang-san jumped onto the table."
- (23) Zhang-san zai zhuozi-shang tiao.  
 Z. at table-on jump  
 "Zhang-san is jumping (up and down) on the table."

Another aspect that emerges when comparing the meaning of preverbal PPs with that of post-verbal ones is that, while the former can be added to any verb, the latter is found only with certain verbs. According to Travis, this difference derives from the fact that post-verbal PPs actually contribute to the verb's meaning as they are subcategorized by the verbal head itself and, therefore, they «are part of the verb's argument structure and get their  $\theta$ -role from the verb» (Travis 1984, p. 50). With preverbal PPs, on the other hand,  $\theta$ -roles are assigned to the object NP by the prepositions:

Let us say, then, that the difference between postverbal and preverbal PPs is that the preverbal PPs are thematically independent of the verb, i.e., do not get  $\theta$ -marking by the V. The NP within the PP is assigned its  $\theta$ -role by the preposition. As shown above, the post-verbal PPs are not independent of the V. they are arguments of the verb and get their  $\theta$ -marking, at least in part, from the verb. (Travis 1984, pp. 52-53).

Second, while MM1 is basically VO, an object complement can sometimes precede the verb. Building on an observation made by Li & Thompson (1975), Travis points out that that «the object of the verb [can] be placed preverbally if it is preceded by what L&T call an object marker, *ba*», as shown in (24) (Travis 1984, p. 42):

- (24) haizi ba shu mai le. (Li & Thompson 1975)  
 child ba book buy ASP  
 “The child bought the book.”

Focusing now on the change from AC to MM1 with respect to the position of PPs, the fact that in AC both the subcategorized PP (PP1) and the non-subcategorized PP (PP2) are always placed after the verb while the linear order of MM1 is S-PP2-V-O-PP1 shows that «the PP that is not dependent on the verb is now to the verb’s left» (Travis 1984, p. 53). On the one hand, the change undergone by PP2s can be accounted for by speculating that, while AC’s prepositions *gei* and *zai* were co-verbs, in MM1 these same elements have been reanalyzed as prepositions, with their position before the verb reflecting MM1’s change from head-initial to head-final. As far as this parametric reconfiguration is concerned, some independent evidence in this respect can be found by observing the position of aspect. In fact, since aspectual particles in MM are realized on the INFL node, the fact that in sentences such as (20) and (21) (repeated below as 25-26) these elements appear to the right of the verb further supports the idea that MM is a head-final language (cf. *ibidem*).

- (25) ta gei wo mai le chezi le.  
 he for me sell asp car ASP  
 (26) ta mai gei wo chezi le.  
 he sell to me car ASP

On the other hand, the fact that PP1s have remained in post-verbal position despite the overall change from head-initial to head-final is due to the existence, according to Travis, of an additional parameter accounting for the direction of theta-role assignment. This parameter, which is assumed to be independent from the one accounting for head-directionality, is set in such a way that in MM1 both prepositions and verbs must appear to the left of their argument NPs in order to theta-mark them:

We propose that the direction of  $\theta$ -role assignment is another parameter which determines word order in language. We can then claim that while MM1 is head final, it assigns  $\theta$ -roles to the right. (Travis 1984, pp. 54-55):

At this point, however, there is still one issue which seems to invalidate the above proposal. More precisely, although the idea behind rightward theta-role assignment implies that only non-arguments of the verb may occur to the verb's left, this does not seem to be the case for the NP argument of the object marker *ba* in so-called *ba*-constructions seen above in (26), which occurs in preverbal position. Nonetheless, a clue as to how to solve the problem is given us by observing passive constructions in Chinese. As pointed out by the author, in this language «passive is created [...] by having the d-structure subject in a *bei* NP ("by NP") construction», as shown in (27) (Travis 1984, p. 54):

- (27) Neizhi ma    bei ta    qi    de hen lei.  
       That    horse    by    him    ride    till    very    tired  
       "That horse was ridden by him till it got very tired."

Assuming the NP appearing to the left of the verb in (27) to be the verb's external argument, the solution put forth by Travis is that *ba* and *bei* share the property of absorbing a  $\theta$ -role of the verb – with *ba* absorbing [patient] and *bei* absorbing [agent] – and, subsequently, of assigning it to the respective NP complements following them. In these terms, therefore, the fact that «objects have a choice of either being assigned a  $\theta$ -role directly by the verb [...] or by the object marker *ba* in a pre-verbal position» explains why object NPs can occur both before and after the verb without running counter to rightward theta-role assignment (Travis 1984, p. 55).

To summarize so far, while in AC linear word order can straightforwardly be accounted for by assuming that the head-complements parameter is specified differently for VPs and NPs, Travis's analysis argues that this parameter is not the only one involved in MM's linear order. While being specified as head-final, in fact, in MM the effects of this parameter setting are, in certain specific syntactic contexts, overridden by a further parameter determining the direction of theta role assignment, which is in turn specified as rightwards.

The last step of Travis's analysis of Chinese is to analyze a projected stage of Modern Mandarin which she refers to as MM2 (cf. Travis 1984, p. 39). The first crucial aspect of this synchronic system is that, as far as verbal complements are concerned, datives and other PP arguments appear to the left of the verb, with only bare objects occurring in postverbal position. Given what has been stated above about the property of the *ba*-construction, which in turn allows the preverbal NP object to be theta-marked by its



preceding object marker instead of the verbal head, Travis assumes that, in MM2, «the independent role assignment properties of *ba* are lost, and that the verb, at least compositionally, also assigns the  $\theta$ -role to the object of *ba*» (Travis 1984, p. 56). In this sense, while MM2 is still head-final as MM1, the fact that «all the complements of the verb appear to its left except for the element which requires case, i.e., the object» suggests not only that theta-roles are now being assigned to the left but, furthermore, that MM2 assigns Case rightwards (Travis 1984, p. 57). Accordingly, the set of parameters proposed by Travis now includes head-directionality, the theta-role parameter and the Case parameter:

I propose here another parameter which accounts for this synchronic description. This parameter is used to describe a language which has every element of the verb phrase preverbally, except for the “bare” object, which appears post-verbally. (Travis 1984, pp. 56-57)

As far as other categories are concerned, further observations made by the author confirm what has been advanced with respect to the internal structure of VP. First, NPs are head-final, which is expected given the fact that N is not a Case assigner. Second, since the opposite is true for P, it is predicted that PPs – more precisely, those PPs whose head is a Case assigner – are head-initial, and this is shown to be the case. (cf. Travis 1984, p. 58).

Crucially, one important aspect which is not *directly* accounted for by the syntactic parameters reviewed so far is the syntactic realization of the verb’s subject. In fact, if on the one hand «every complement position in the  $\theta$ -grid corresponds directly to a complement position in the strict subcategorization frame» (Stowell 1981, p. 35), on the other hand «a subject is never a complement as it is never in the subcategorization frame of a lexical item» (Travis 1984, p. 31). The reason for this is that, following Williams (1981), the domain for subcategorization is limited to the maximal projection containing the verbal head and its complements (hence regarded as *internal arguments*), while the subject is never strictly subcategorized as it is generated outside its verb’s maximal projection (and thus corresponds to the *external argument*). This, in turn, implicates that a subject cannot possibly receive its theta-role directly from the verb although the Theta-Criterion forces it to be theta-marked. Hence, in the technical sense, the external theta-role has to be assigned by the verb *indirectly*, as this is the only possible way to maintain the initial assumption that the thematic grid is encoded in the verb’s lexical entry. As noted by Chomsky (1981a):

In case [ $\beta$  is the subject of  $\alpha$ ], if  $\alpha$  marks  $\beta$  (a position or category), let us say that the lexical head of  $\alpha$  *indirectly  $\theta$ -marks*  $\beta$ . Direct or indirect  $\theta$ -marking are therefore properties of lexical items determined by the lexicon. If  $\alpha$  directly or indirectly  $\theta$ -marks  $\beta$ , we say that  $\alpha$  *selects*  $\beta$ . A verb, for example, selects its complements and also selects its subject if it participates in assigning a  $\theta$ -role to the subject. (Chomsky 1993 [1981a], p. 38)

Regarding the question of the syntactic realization of the verb's subject, in Chomsky (1981a) a solution is proposed in order to force subjects to be inserted when obligatorily required to do so. This solution, which was initially formulated as the *principle P*, would later be known as the Extended Projection Principle (EPP) and corresponds to «the structural requirement that certain configurations» – namely, clauses – «must have subjects» (Chomsky 1993 [1981a], p. 27):

Obligatory insertion of the NP [...] follows from the fact that the constructions illustrated require subjects for some structural reason; call it the principle *P*. Clearly, *P* does not derive from  $\theta$ -theory; [...] Nor does *P* derive from considerations of subcategorization. Verbs do not subcategorize for subjects, which may be freely missing when *P* is inapplicable [...]. (Chomsky 1993 [1981a], p. 26)

With the EPP taking care of the external argument's insertion, the fundamental asymmetry between subject and non-subject theta-marking was dealt by GB-Theory by assuming that, while «an object NP [...] receives its internal  $\theta$ -role from the head of which it is a complement», «the subject NP receives its  $\theta$ -role compositionally [...] from the maximal projection of VP», in the sense that its semantic status is determined not only by the verb but, indirectly, by the entire predicate (Travis 1984, p. 91):

We will say that  $\alpha$   *$\theta$ -marks the category  $\beta$*  if  $\alpha$   $\theta$ -marks the position occupied by  $\beta$  or a trace of  $\beta$ . [...] This account presupposes that the  $\theta$ -role of a subject (where it has one) is determined by the VP of S rather than by the verbal head of this VP. (Chomsky 1993 [1981a], p. 37)

Considering the existence of such nontrivial differences between internal arguments and the subject NP, Travis's hypothesis in this respect is that «the direction of predication may be different from the direction of direct  $\theta$ -marking» (Travis 1984, p. 92). Crucially, this difference seems exactly to be the factor accounting for the fact that, in Chinese but also in languages such as English, the subject is placed before the verb while the internal arguments are all realized on the other side of the verb (*ibidem*):

- (28) ta mai gei wo chezi le.  
 he sell to me car INFL  
 “He sold me a car.”

In parametric terms, therefore, the syntactic model put forth in Travis (1984) in order to derive the word order patterns of Archaic Chinese and the two varieties of Modern Mandarin briefly presented above is based on four binary parameters, the latter of which is the one accounting for the direction of predication/assignment of the external theta-role.

As compared with the parametric proposal suggested in Stowell (1981), the theory of linearization put forth by Travis goes considerably deeper in its attempt to characterize parametric variation. In addition to the number of syntactic parameters involved in her account, the strength of Travis’s proposal lies, first of all, in the systematization of a markedness-based model which is similar, in certain aspects, to the one that would be proposed by Manzini & Wexler (1987) in connection with the parametrization of the notion of governing category. Focusing on the linear distribution of verbal complements, Travis proposes that, considering the subdomain parameters accounting for the direction of theta-role assignment and Case assignment, «only one parameter may be set outside of the default case», the latter being meant as the specific option specified with respect to head directionality (Travis 1984, p. 104). The assumed rationale behind this stipulation concerns the implicational relationships between the head-complement parameter and the particular subdomains of grammar involved in  $\theta$ -role assignment and Case assignment, whose settings in the languages seen so far is summarized in the table shown below (*ibidem*):

(29)

	word order	headedness	$\theta$ -roles	case
AC:	V O PP1 PP2	initial	——	——
MM1:	PP2 V O PP1	final	right	——
MM2:	PP2 PP1 V O	final	——	right

Assuming we have a language in which, as exemplified by AC, the default case corresponds to “initial headness”, that is, “X’  $\rightarrow$  X ...”, setting  $\theta$ -role assignment as rightwards would yield vacuous results since the latter’s effects would be indistinguishable from the ones emerging from initial headness. Accordingly, the same applies to the

direction of Case-marking, as also the possible configurations predicted by this parameter are a subset of the ones predicted by the head-complement parameter. Ultimately, then, whenever one of these more specific/less predictive parameters needs to be specified, all the internal arguments whose position has not been accounted for in this way will necessarily fall within the scope of the minimally specific/maximally predictive schema provided by the head-complement, which, in turn, will place them on the opposite side of the verb. (cf. Travis 1984, p. 105). Accordingly, as noted by Travis:

[...] if the direction of case assignment must be specified, then the direction of  $\theta$ -assignment may not be specified, and vice versa. (Travis 1984, pp. 104-105)

The second innovative aspect of Travis's proposal is that it also formulates an explicit distinction concerning the structural level on which the parameters so far discussed take effect. Starting from the classic GB assumption that «d-structure is the “pure representation of GF- $\theta$ ”», the direction of theta-role assignment is implied to have an effect on D-Structure, while leaving S-structure unaffected (Travis 1984, p. 105). Conversely, the direction of Case-assignment would impose the effects of its potential setting only at S-Structure, but not at D-Structure. It is in light of these considerations that the author proposes the following D-Structures schemata for AC, MM1 and MM2 (Travis 1984, p. 106):

- (30) a. Archaic Chinese:  
[ V NP PP1 PP2 ]
- b. MM1:  
[ PP2 V NP PP1 ]
- c. MM2:  
[ PP2 PP1 NP V ]

What is particularly interesting about (30) is that, assuming that MM2 displays the based-generated word order PP2-PP1-NP-V and the surface order PP2 PP1 V O, it emerges that it is rightward Case-assignment which drives the overt movement of the bare NP object from preverbal to postverbal position. From a P&P perspective, therefore, although in GB Theory linearization in the literal sense is still assumed to be the consequence of the parametrization of the X-bar module of UG as envisioned in Graffi (1980), the approach followed by Travis showed not only that linear order of the verb with respect to its arguments was a result of the interaction between various subsystems of UG

(in the matter in question, X-bar theory, theta-theory and Case theory), but also that, in the spirit of Stowell (1981), its various cross-linguistic variation patterns could be accounted for without the need to resort to language-particular phrase-structure rules.

In conclusion, Travis (1984) marked a turning point in the development of the head-complement parameter. Although based on the now long abandoned assumption that each subsystems of UG potentially admits a certain degree of parametrization, its proposal that headedness, theta-marking and Case assignment might potentially be opposite in their directionality and, not less importantly, that there exists an implicational relation among a series of parametric values qualifies Travis's theory of linearization as a precursor of today's microparametric inquiries.

## 5.2 – The crisis of the head-complement parameter

With the shift from GB theory to the MP, the idea that parametric variation should be identified with variation in the lexical properties of functional heads definitely started to have edge over traditional macroparameters. Consequently, although having the merit of addressing linear order from a P&P perspective, it was only a matter of time before macroparametric accounts such as that put forth by Travis (1984) would be abandoned in favor of a model which would not treat linguistic variation in terms of parametrization of UG. In these terms, the traditional head-complement parameter would be definitely supplanted by the theory proposed in Kayne's (1994) book *The Antisymmetry of Syntax*.

### 5.2.1 – *The antisymmetry of syntax*

Contrary to all the works which have been reviewed so far in this chapter, the leading idea argued for by Kayne (1994) is that linear order is not subject to variation as «the human language faculty is in fact rigidly inflexible when it comes to the relation between hierarchical structure and linear order» (Kayne 1994, p. xiii). The process through which Kayne arrives at this hypothesis does not start, as has usually been done, from considering some abstract universal principle regarding phrase structure (as X-bar theory in the case of Graffi (1980), Stowell (1981), and Travis (1984)), but from the observation of what the author indicates as the defining properties of linear order. These properties are essentially three (Kayne 1994, p. 4):

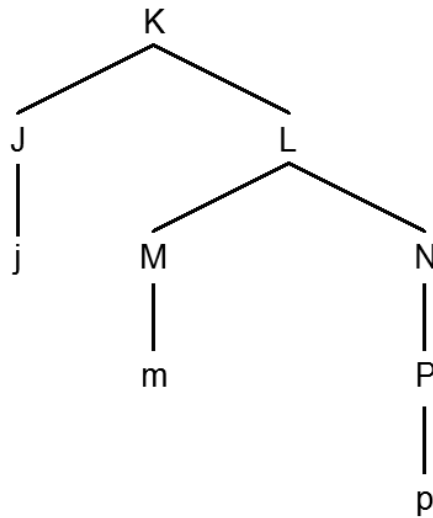
- (31)
- a. It is transitive; that is,  $xLy \ \& \ yLz \rightarrow xLz$ .
  - b. It is total; that is, it must cover all the members of the set: for all distinct  $x, y$ , either  $xLy$  or  $yLx$ .
  - c. It is antisymmetric, that is, not  $(xLy \ \& \ yLx)$ .

Looking at phrase structure, there is a specific aspect of this system which strongly resembles linear order in its properties, namely, the dominance relation on nonterminal nodes. Analogously to linear order, the dominance relation is both transitive and antisymmetric, although not total. This latter aspect is evident when considering the sisterhood relation intercurring between two nodes symmetrically c-commanding each other. However, while not total in itself, as noted by the author the dominance relation becomes total provided it is localized, that is, «when restricted to the set of nodes dominating a given node» (Kayne 1994, p. 4). Considering now the relation of asymmetric c-command between nonterminal nodes, which in turn can be seen as an instantiation of localized dominance, and assuming a strictly binary branching structure as that derived from X-bar theory, the author's intuition is that «there should be a very close match between the linear ordering relation on the set of terminals and some comparable relation on nonterminals», with *comparable* here meant as *locally linear* (Kayne 1994, p. 5). This is precisely what happens in the asymmetric c-command relation: in fact, given two nonterminal nodes  $X$  and  $Y$  such as  $X$  asymmetrically c-commands  $Y$ , if there is a third nonterminal node  $Z$  which asymmetrically c-commands  $Y$ , then either  $Z$  asymmetrically c-commands  $X$  or  $X$  asymmetrically c-commands  $Z$ . According to Kayne, the locally linear relation intercurring between  $Z$  and  $X$  is the same one intercurring between two terminal nodes which are mapped linearly after each other. Therefore, the author concludes that «asymmetric c-command is closely matched to the linear order of terminals», as formulated in his Linear Correspondence Axiom (LCA) (*ibidem*):

- (32) Linear Correspondence Axiom:  
 $d(A)$  is a linear ordering of  $T$ .  
 (Kayne 1994, p. 6)

In (32),  $d$  stands for the dominance relation intercurring between nonterminal nodes and terminal nodes,  $A$  is the set of all pairs of nonterminal nodes, and  $T$  is the set of terminals. The way in which the LCA works is exemplified by Kayne with respect to the phrase marker in (33):

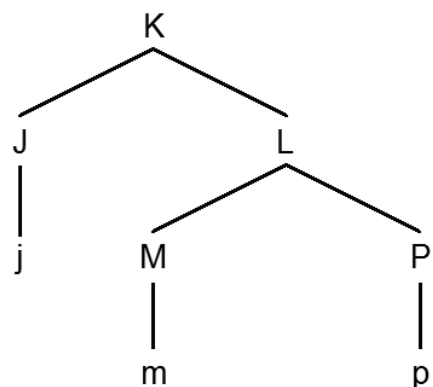
(33)



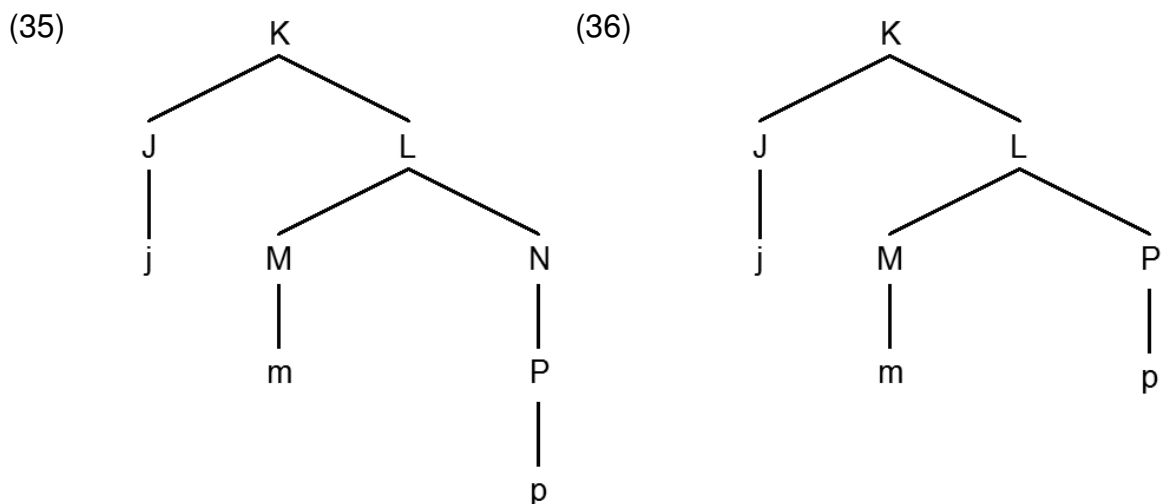
In (33), J asymmetrically c-commands M, N and P, while M asymmetrically c-commands P. The algorithm provided by the LCA maps the asymmetric c-command relations intercurring between the nonterminal pairs  $\langle J, M \rangle$ ,  $\langle J, N \rangle$ ,  $\langle J, P \rangle$ ,  $\langle M, P \rangle$  onto the terminal pairs  $\langle j, m \rangle$ ,  $\langle j, p \rangle$ ,  $\langle m, p \rangle$  which, being instances of linear order, do conform to the properties listed in (31). First, they are transitive: J precedes M and M precedes P, hence J transitively precedes P. Second, they are total, as the resulting set of terminals T covers the linear order of all terminal nodes in a given phrase marker. Third, they are antisymmetric, as any of these precedence relations cannot be set the other way around. Accordingly,  $\langle j, m \rangle$ ,  $\langle j, p \rangle$ ,  $\langle m, p \rangle$  constitute the linear ordering  $\{j, m, p\}$ .

If on the one hand Kayne's theory of linearization represents, as shown with respect to the phrase marker in (33), an algorithm mapping asymmetric c-command onto linear precedence, on the other hand this model has far deeper theoretical implications than it seems at first sight because, as the author himself suggests, «linear order turns out to be more fundamental to syntax than is normally thought» (Kayne 1994, p. 131). These implications can in turn be grasped more easily by looking at the phrase marker shown below:

(34)



As easily noticeable, the difference between (33) and (34) consists in the latter's lack of an intermediate node between nonterminal P and terminal p, which results in the fact that the heads taken as input by the algorithm in (32) are J, M and P. However, if on the one hand J asymmetrically c-commands both M and P hence giving  $\langle J, M \rangle$ ,  $\langle J, P \rangle$ , on the other hand there is no asymmetric c-command between M and P. Since M and P cannot be ordered with respect to each other, neither can their respective terminal nodes. Consequently, as  $\{j, m, p\}$  fails to be total in the sense of (31b), «[34] fails to meet the requirement imposed by the LCA and is therefore not an admissible phrase marker» (Kayne 1994, p. 8). In these terms, the first implication of Kayne's model is that any syntactic representation violating the LCA cannot in principle be generated by narrow syntax as they cannot be linearized. Taking (33) and (34) to respectively correspond to (35) and (36):

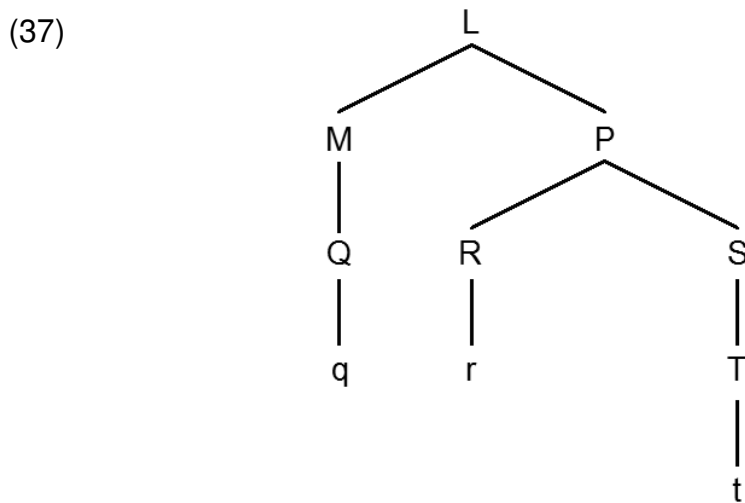


the ban on (34-36) can be accounted for by saying that, as axiomatically assumed by X-bar theory, the complement of a head cannot itself be a head, and hence deriving the more general restriction that «a phrase cannot have two heads» (*ibidem*).

As acknowledged by the author himself, one further refinement needed by this working hypothesis concerns the notion of specifier. According to Kayne's terminology, «a nonterminal that dominates no other nonterminal [is] a *head*» while «a nonterminal that does dominate at least one other nonterminal will be a *nonhead*», with only the former typology constituting the set of appropriate arguments  $A$  for the function represented by  $d$  (Kayne 1994, p. 11). By considering, therefore, only those nodes which the author refers to as heads, if on the one hand the basic clausal representation (35) is allowed in terms of



LCA, on the other hand the specifier position of VP cannot host a head node since, analogously to the complement position, it should instead host an XP as shown in (37):



As far as antisymmetry goes, though, in this case the remedy seems worse than the disease; in fact, now the asymmetric c-command set  $A$  for (37) is  $\langle M, R \rangle$ ,  $\langle M, S \rangle$ ,  $\langle M, T \rangle$ ,  $\langle R, T \rangle$ ,  $\langle P, Q \rangle$ , with  $P$  asymmetrically c-commanding  $Q$ . Since the corresponding linear mappings of  $\langle M, R \rangle$  and  $\langle P, Q \rangle$  are, respectively,  $\langle q, r \rangle$  and  $\langle r, q \rangle$ , the antisymmetry requirement stated in (31c) is not met, hence the impossibility to correctly linearize (37). In order to solve this problem, Kayne's proposal is to modify the definition of c-command: according, on the one hand, to the distinction between *categories* and *segments* which was originally introduced by May (1985) and adopted by Chomsky (1986a); and on the other hand, to the notion of *exclusion* (Chomsky (1986a)). Regarding the distinction between category and segment, Chomsky (1986a) proposes the following:

[...] in a structure of the form [38], a typical adjunction structure with  $\alpha$  adjoined to  $\beta$ ,  $\alpha$  is not dominated by the category  $\beta$ ; rather,  $\beta$  consists of two "segments," and a category is dominated by  $\beta$  only if it is dominated by both of these segments. (Chomsky 1986a, p. 7)

(38)  $[_\beta \alpha [_\beta \dots]]$

Following Chomsky's proposal, if in (37) the node  $M$  were *adjoined* to  $P$  instead of *merged* to  $P$ ,  $M$  would not be dominated by the resulting higher node  $P$  as the latter would not be a category, but only one of the two segments forming  $P$ .

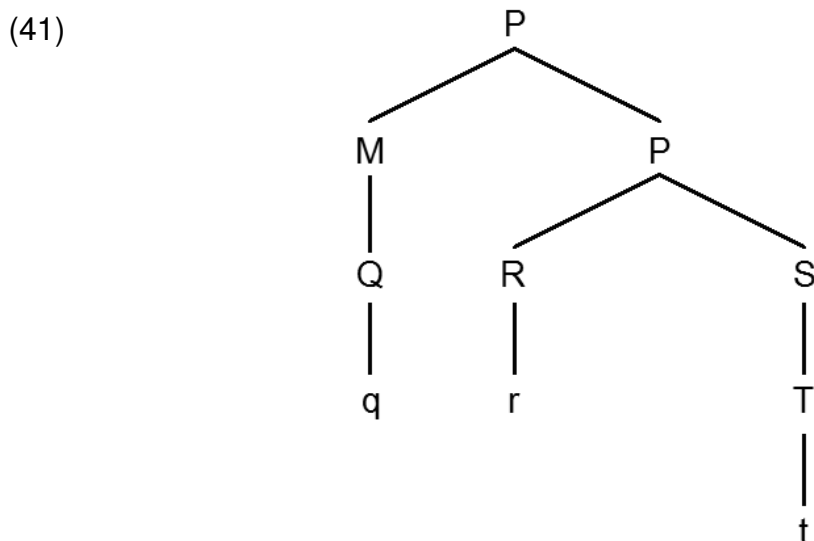
Second, exclusion is defined by Chomsky (1986a) as in (39) (Chomsky 1986a, p. 9):

(39)  $\alpha$  excludes  $\beta$  if no segment of  $\alpha$  dominates  $\beta$ .

Building on these formulations, Kayne restricts the relation of c-command to categories by stipulating that «a segment cannot enter into a c-command relation» (Kayne 1994, p. 16):

(40)  $X$  c-commands  $Y$  iff  $X$  and  $Y$  are categories and  $X$  excludes  $Y$  and every category that dominates  $X$  dominates  $Y$ .

Considering in this light that «specifiers are a case of adjunction» (Kayne 1994, p. 22), then the counterpart of (37) would be (41):



As pointed out by the author, while in (37) the problem was that  $P$  asymmetrically c-commanded  $Q$ , in (41) lower  $P$  cannot c-command into the domain of  $M$  because lower  $P$  is a segment and not a category. Consequently, as there is no way for the  $d(A)$  of (41) to give as is output  $\langle r, q \rangle$ , the antisymmetry requirement stated in (31c) is not violated and the terminal nodes  $q$ ,  $r$  and  $t$  are linearized correctly.

Summing up so far, according to Kayne's antisymmetric model, the exclusion of any potential unsuitable phrase marker from the derivation follows directly from the lack of antisymmetry in any two of the totality of its head terms. In these terms, not only do the principles of X-bar theory derive as theorems from the LCA, but also linear order may be regarded in this light as an interface condition forcing syntactic representation to conform to those properties which are required by a given set of terminal nodes in order to be linearly ordered:

As a result of the LCA, the property of antisymmetry that a linear ordering has is inherited by the hierarchical structure. I have argued that this is behind X-bar theory, or rather, that X-bar theory, although largely accurate in its standard form, should not be considered to be a primitive part of syntactic theory (i.e., of UG). What is primitive in UG is the LCA, from which follow familiar X-bar-theoretic properties [...]. Combined with a fairly standard definition of c-command in terms of category (as opposed to segment), the LCA goes beyond X-bar theory in the extent to which it limits phrase structure diversity. (Kayne 1994, p. 131)

This is only the more general implication of the theory reviewed here. Starting from the assumption that, given the pressure exerted by the LCA on syntactic derivation, the specifier must always asymmetrically c-command the head and the head must always do the same with the complement, Kayne concludes that, in principle, the only possible constituent orders are Specifier-Head-Complement (S-H-C) and Complement-Head-Specifier (C-H-S), with all other permutations being excluded by the requirement that «specifier and complement be on opposite sides of the head» (Kayne 1994, p. 35):

More generally put, no matter how complex the specifier or complement, it will always be the case, in any phrase marker, that specifier and complement are on opposite sides of the head. In other words, if we represent head, specifier, and complement as H, S, and C, then the conclusion so far is that of the six permutations of H, S and C, only two are permitted by the theory, namely, S-H-C and C-H-S. (*ibidem*)

At this point, Kayne continues, «of the two orders, the former is a significantly more plausible universal than is the latter» (*ibidem*). While also considering empirical evidence, especially with respect to the overall prevalence of the specifier-head order higher across relevant cross-linguistic data, Kayne argues for the idea that S-H-C is the only order available mainly on theoretical grounds. His proposal is the following: considering a linear ordering  $d(A)$  consisting of all the terminals in a given phrase marker, the fact that every asymmetrical relation  $\langle x, y \rangle$  intercurring between any of the pairs contained in  $d(A)$  is interpreted as “x precedes y” rather than as “x follows y” is «ultimately related to the asymmetry of time» (Kayne 1994, p. 38). This idea is implemented by postulating the existence, among the nodes of every phrase marker, of an abstract node A adjoined to the root node and asymmetrically c-commanding every other node from an adjoined position. Being no different from every other node in the phrase marker in no other respect but its abstractness, Kayne specifies not only that «A should be taken to dominate a terminal element», but also that this latter terminal element could in principle either precede or

follow all the other terminals (Kayne 1994, p. 37). Between the two admissible choices, Kayne opts for the former. In fact, assuming the string of terminals «as being associated with a string of time slots», the author further postulates that «what is paired with each time slot is not simply the corresponding terminal, but the substring of terminals ending with that terminal (i.e., the substring produced up to that time)», as shown in (42) (*ibidem*):

(42) a, ab, abc, abcd, abcdz

In (42), a and z each represents one possible position of the abstract terminal element dominated by the abstract nonterminal A. Given the fact that «that root node for terminals must be in some fixed relation to every terminal in every substring» as its nonterminal counterpart A, according to the author this implies that the abstract terminal for A is a and not z, as only the former is always present in each single time slot (*ibidem*).

At this point, it still remains to be explained «whether  $\langle x, y \rangle$  is “x precedes y” or “x follows y”», that is, whether asymmetric c-command is mapped onto linear precedence or subsequence (*ibidem*). Starting from the assumption that, as stated above, A asymmetrically c-command every other node in the phrase marker, the same must also be true for a with respect to every other terminal node. Then, as a is the beginning terminal in (42), according to Kayne it must follow that asymmetric c-command is mapped onto linear precedence, and not the other way around.

Returning now to the implications of Kayne’s theory, if on the one hand the LCA poses a number of restrictions on the possible phrase structures, on the other hand this hypothesis has also crucial consequences for crosslinguistic word-order variation. In fact, if the only linear order made available by UG corresponds to S-H-C, then the existence of the head-complement parameter of Stowell (1981) and Travis (1984) (or, more in general, of directionality parameters as the ones envisioned in Graffi (1980)), is in principle excluded from an account of cross-linguistic variation, with other observed orders being the result of movement operations further manipulating syntactic structures:

If UG unfailingly imposes S-H-C order, there cannot be any directionality parameter in the standard sense of the term. The difference between so-called head-initial languages and so-called head-final languages cannot be due to a parametric setting whereby complement positions in the latter type precede their associated heads. Instead, we must think of word order variation in terms of different combinations of movements. Note first that from the present perspective, any movement of a phrase upward to a c-commanding position must be

leftward. This is so, for the simple reason that asymmetric c-command implies precedence (Kayne 1994, p. 47).

Although Kayne's (1994) remaining chapters provide extensive empirical analyses supporting this hypothesis, what is more relevant to the aim of the present thesis is that, with this seminal work, Kayne effectively stripped the traditional head-directionality (macro)parameter of its explicit theoretical status, hence turning it into a mere descriptive device. Moreover, the fact that the LCA provided a series of principled constraints on phrase structure added to its importance within the Minimalist framework which was beginning to take shape precisely in these years.

### 5.3 – Some recent proposals regarding head-directionality

As well expressed by Abels & Neeleman (2007), «Kayne (1994) was instrumental in putting linear asymmetries on the generative research agenda» (Abels & Neeleman 2007, p. 1). In agreement with the parallel MP's effort to reduce both methodological and computational complexity, the LCA had the effect not only of eliminating X-bar theory from the set of once assumed subsystems of UG, but also of dispensing with the head complement parameter and its underlying assumption of the possible existence of a distinct underlying linear order for each possible language. As Kayne (1994) himself recognized:

I have derived the result that specifier-head-complement order is the only order made available by UG and consequently that there can be no directionality parameter for word order. (Kayne 1994, p. 132)

However, although the LCA is still acknowledged as an appealing tool for dealing with the issue of cross-linguistic variation in word order, its role has undergone a profound change since the time of its formulation, the main reason being that the conception of a totally asymmetrical syntax which was upheld in Kayne (1994) has eventually clashed with later developments in generative theory. In these terms, the main problematic aspect in this respect concerns the compatibility of Kayne's theory of linearization with some of the core assumptions of Chomsky's (1995a,b) bare phrase structure theory (BPS) (cf. Chomsky 1995a, p. 336).

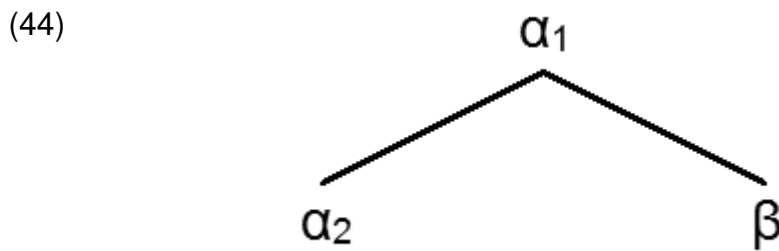
BPS represents a still ongoing methodological approach whereby, analogously to any other theoretical construct of Generative Grammar, phrase structure theory is being evaluated against what Chomsky refers to as *bare output conditions*, that is, «the minimal need that linguistic computations must satisfy [...] to connect interface representations» (Chomsky 2002, p. 41). As far as traditional X-bar theory is concerned, Chomsky's main argument is that, under such minimalist conception, «phrase structure representation is “bare”, excluding anything beyond lexical features and objects constructed from them» (Chomsky 1995a, p. 245). Starting from the assumption that «at the LF [Logical Form] interface, it must be possible to access» at least «a lexical item LI and its nonphonological properties LF(LI)» and «some larger units constructed of lexical items» (Chomsky 1995a, p. 242), the only syntactic objects which narrow syntax cannot arguably dispense with are, according to Chomsky, lexical items on the one hand, and the larger units created by the operation Merge on the other (Chomsky 1995a, p. 243):

- (43) a. lexical items  
 b.  $K = \{ \gamma, \{ \alpha, \beta \} \}$ , where  $\alpha, \beta$  are objects and  $\gamma$  is the label of  $K$

The role of forming larger units out of two objects  $\alpha$  and  $\beta$  – each of which can in turn correspond to either a lexical item or a syntactic object already formed – is undertaken by the basic computational operation: Merge. As shown in (43b), Merge has the property of defining certain basic relations between phrasal constituents. First, «applied to two objects  $\alpha$  and  $\beta$ , Merge forms the new object  $K$ , eliminating  $\alpha$  and  $\beta$ » (*ibidem*). Although  $\alpha$  and  $\beta$  are the constituents of  $K$ , to specify  $K$  simply as the set  $\{ \alpha, \beta \}$  would not suffice to meet bare output conditions; in fact, as Chomsky notes, «verbal and nominal elements are interpreted differently at LF and behave differently in the phonological component» (*ibidem*). In these terms, what  $K$  needs is to be specified with a *label*. Hence we have  $K = \{ \gamma, \{ \alpha, \beta \} \}$ , where « $\gamma$  identifies the type to which  $K$  belongs, indicating its relevant properties» (*ibidem*).

Second, the role of Merge also encompasses labelling. According to Chomsky, «the label of  $K$  is determined derivationally [...] rather than being derived representationally at some later stage of the derivation» (*ibidem*). Stripping down phrase structure theory to the bare minimum, Chomsky's proposal in this respect is that category labels in the spirit of X-bar theory are not necessary. More precisely, the label  $\gamma$  of a phrase  $K$  is nothing but the head of  $K$ , which in turn necessarily corresponds to either  $\alpha$  or  $\beta$ . Therefore, assuming for

the sake of convenience that the head of  $K$  is  $\alpha$ ,  $\alpha$  further projects, thus giving  $K = \{ \alpha, \{ \alpha, \beta \} \}$ , as shown in (44) (Chomsky 1995a, pp. 244-245):



What distinguishes a phrase structure representation such as (44) from its x-bar counterpart is the fact that, under this bare conception of phrase structure, «there is no such thing as a nonbranching projection» (Chomsky 1995a, pp. 246). Moreover, what is notably absent in Chomsky's BPS compared to the X-bar schema is also a specification of the difference between maximal ( $X^{\max}$ ) and minimal ( $X^{\min}$ ) projections in the structures generated by narrow syntax (cf. Chomsky 1995a, p. 242). On the one hand, the former property derives from the very nature of Merge: being derivationally asymmetric, Merge «project[s] one of the objects to which it applies, its head becoming the label of the complex formed», hence the impossibility for any other projection, including a vacuous, non-branching one, to be generated instead of the one corresponding to the phrase's label (Chomsky 1995b, p. 63). On the other hand, the latter property is based, according to Chomsky, on the purely relational status of the distinction between minimal and maximal projections. There is, in fact, no other principled difference between categories which can justify such an arbitrary division apart from the fact that one further projects, while the other does not:

Given a phrase marker, a category that does not project any further is a maximal projection  $XP$  and one that is not a projection at all is a minimal projection  $X^0$ ; any other is an  $X'$ , invisible at the interface and for computation. (Chomsky 1995b, p. 61)

Returning now to Kayne's theory of linearization, in Chomsky (1995a) it is suggested that the crucial factor undermining a "strong" version of the LCA conceived as an constraint on both linear order and syntactic derivation is the fact that, while on the one hand «certain stipulated properties of X-bar theory can be derived from the LCA», on the other hand «the derivation of these properties relies crucially not just on the LCA, but on features of standard X-bar theory that are abandoned in the bare theory» (Chomsky

1995a, pp. 335-336). In particular, given that BPS theory dispenses with non-branching projections, in a bare phrase marker such as (44) the head  $\alpha$  and the complement  $\beta$  are in a mutual c-command relation. Therefore, since no linear order can be determined for this pair of terminals, the structure violated the LCA (cf. Chomsky 1995a, p. 337).

In light of the above facts, Chomsky proposed to maintain a weaker version of the LCA by taking it to be «a principle of the phonological component that applies to the output of Morphology» – that is, a linearization strategy rather than an overarching constraint as assumed in its stronger version (Chomsky 1995a, p. 340). In these terms, the fact that Kayne's (LCA) proposal has been called into question has awakened a new interest in the possibility that the head-complement parameter, although no more compliant with an overspecified view of UG, could be a viable means by which account for the fact that head/complement ordering is not stable either across languages or across the constituents belonging to the same language.

### *5.3.1 – A minimalist reformulation of the head-complement parameter*

As the reconceptualization of the LCA as a linearization strategy left the field open once again for the idea of a parametrization of head directionality, a number of proposals have been made to reconcile the existence of the head-complement parameter with key Minimalist assumptions. With the demise of the macroparametric approach in modern minimalism (Newmeyer 2004, 2005), one possible way to deal with the issue of cross-linguistic variation in linear order has been, in agreement with the view that «ordering is restricted to externalization of internal computation to the sensory-motor system, and plays no role in core syntax and semantics», to regard the head-complement parameter as applying in the mapping to PF rather than in narrow syntax (Berwick & Chomsky 2011, p. 29). Regarding such assumption of a «clear division of labor between hierarchical and linear order» (Berwick & Chomsky 2016, p. 120), this is the view entertained in Marc Richards's paper *Two kinds of variation in a minimalist system* (2008), which further develops and modernizes some of the conclusions drawn in his PhD dissertation *Object Shift and Scrambling in North and West Germanic: A Case Study in Symmetrical Syntax* (2004).

In these works, Richards's aim is to address what he argues to be another potential problem of Kayne's (1994) original proposal and which, according to his analysis, arises even if considering the LCA purely as a linearization strategy. As stated in Richards (2004,



2008), the main problematic aspect in this respect concerns a specific development of Chomsky's BPS theory in modern Minimalism, namely, the distinction between Set-Merge and Pair-Merge (Chomsky 2000, 2001a). On the one hand, Set-Merge corresponds to FL's symmetric structure-building operation which takes two objects  $\alpha$  and  $\beta$  and generates simple unordered sets  $\{\alpha, \beta\}$ :

For structure building, we have so far assumed only the free symmetrical operation Merge, yielding syntactic objects that are sets, all binary: call them *simple*. (Chomsky 2004 [2001a], p. 117)

On the other hand, Pair-Merge is an asymmetric adjunction operation which takes two objects  $\alpha$  and  $\beta$  and yields an ordered pair  $\langle \alpha, \beta \rangle$ :

But it is an empirical fact that there is also an asymmetric operation of adjunction, which takes two objects  $\beta$  and  $\alpha$  and forms the ordered pair  $\langle \alpha, \beta \rangle$ ,  $\alpha$  adjoined to  $\beta$ . (*ibidem*)

Focusing on the basic structure-building operation Set-Merge, even assuming Chomsky's (1995a) weaker version of Kayne's (1994) LCA it is clear that, as noted by Richards, since neither  $\alpha$  or  $\beta$  asymmetrically c-commands the other, no ordering can be assigned to their terminals. It is precisely this deficiency that, the author continues, not only does undermine the LCA's basic assumption that the relation required for establishing precedence is antisymmetry, but also motivates the necessity of a parametric device allowing such symmetrical syntactic representations to be decoded by the phonological component:

This lack of intrinsic ordering between two (set-)merged elements (i.e. sisters) would seem to render some kind of head-complement ordering strategy a virtual "conceptual necessity" given that unordered structures are illegible (and thus illegitimate) objects at PF. (Richards 2004, p. 19)

Arguing for the existence of the head-complement parameter, Richards builds on Epstein *et al.* (1998)'s intuition that, rather than asymmetric c-command, it is «C-command, pure and simple, to be the relation that induces a precedence relation among terminals» (Epstein *et al.* 1998, p. 151). In this respect, what Epstein *et al.* (1998) proposed to solve the problem represented by symmetric Merge-pairs was, first of all, to

revise the original formulation of the LCA by removing Kayne's asymmetry stipulation, so that the following revised version is obtained (Epstein *et al.* 1998, p. 151):

- (45) Linear Correspondence Axiom (revised):  
If X C-commands Y, then the terminals in X precede the terminals in Y.

As pointed out by Richards, however, at this point a further problem arises, although of a different sort. In fact, if on the one hand an antisymmetry-driven LCA does not provide the phonological component with sufficient linearization instructions, on the other hand, assuming simple c-command to be responsible for the process of linearization, «since Merge-pairs (such as head-complement) mutually c-command each other, they overdetermine linearization, providing contradictory instructions to PF such that each sister must precede the other» (Richards 2008, p. 148). This apparent paradox did not go unnoticed even by Epstein *et al.*, who proposed a solution in the form of a disambiguating strategy which they aptly called *Precedence Resolution Principle* (PRP) and formulated as given in (46):

- (46) The Precedence Resolution Principle:  
If two (not necessarily distinct) categories symmetrically C-command each other by virtue of some syntactic operation O, ignore all C-command relations of one of the categories to the terms of the other with respect to establishing precedence via the LCA.  
(Epstein *et al.* 1998, p. 152)

The idea behind Epstein *et al.*'s PRP is that, given that syntactic symmetry between two merged categories is invariably mapped into mutual precedence, the linearization component systematically ignores either the Comp > Head or the Head > Comp relation. If the former option is chosen, then a head-initial order results; if the latter option is chosen, then the result is a head-final order. This is illustrated by Richards in the following example, which accounts for the head-complement relation in the VP (Richards 2008, p. 149):

- (47)
- |  |  |                                       |  |
|--|--|---------------------------------------|--|
|  | <i>C-command relations</i><br>→ { V > DP, DP > V } | <i>Ignore</i><br>→ V > DP<br>→ DP > V | <i>PF-order</i><br>→ DP > V (= OV)<br>→ V > DP (=VO) |
|--|--|---------------------------------------|--|

In addition to the configuration exemplified by (47), which corresponds to an instantiation of External Merge, mutual c-command relations are also created by Move/Internal Merge. In this respect, a major difference between Epstein *et al.* (1998) and Richards (2004, 2008) lies in their respective treatments of Internal Merge configurations. In fact, if on the one hand Epstein *et al.* (1998) exclude the possibility that the PRP be applied also to instances of External Merge, as they actually regard the systematic triggering of overt movement as the effect of a change in precedence relations, Richards suggests instead that «[46]/[47] should be generalized to hold of internal as well as external Merge, thus imposing VO/OV “shape” on Move and Merge alike» (Richards 2008, p. 149). In these terms, as noted by Richards:

[...] the PRP should be amended so that it deletes a consistent subset of c-command instructions in any given language (rather than allowing the c-command relation  $DP > V$  to always win out in the case of Move, irrespective of base order, the way it does in Epstein *et al.*). (*ibidem*)

Therefore, contrary to Epstein *et al.* (1998), according to Richards (2004, 2008) the PRP represents a parametrized condition which applies throughout the derivation and embraces all mutual c-command relations, regardless of the specific syntactic operations which contributed to their instantiation. Such parametrized condition, which Richards (2008) calls *Parameterised desymmetrisation*, is *de facto* comparable in its effects to the traditional head-complement parameter but, differently from this latter, is located in PF rather than narrow syntax. Its formulation is given in (48) (*ibidem*):

(48) Parametrized desymmetrization:

Given  $\text{Merge}(\alpha, \beta) \rightarrow \{<\alpha, \beta>, <\beta, \alpha>\}$ :

a. Head-initial = Delete all  $\text{Comp} > \text{Head}$  [i.e.  $\{<\alpha, \beta>, <\beta, \alpha>\} \rightarrow \{<\alpha, \beta>\}$ ]

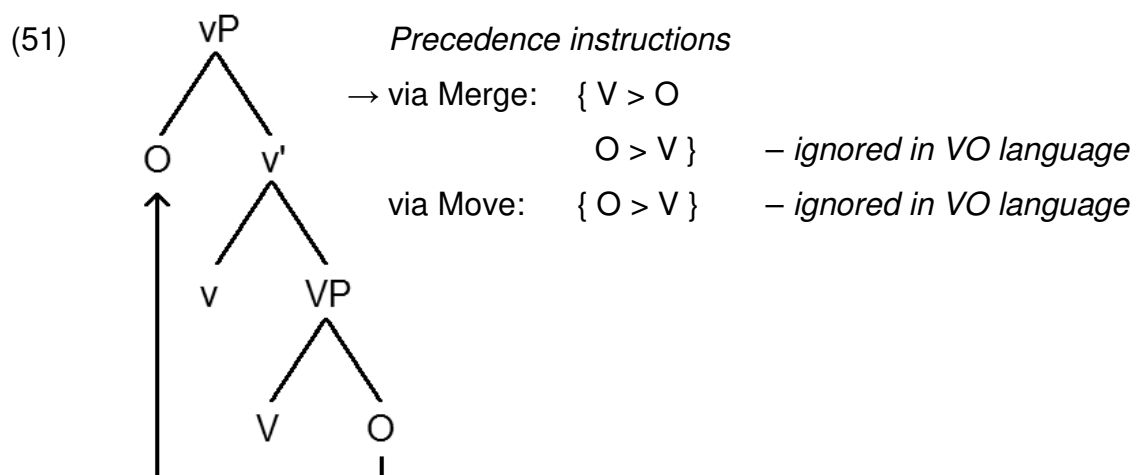
b. Head-final = Delete all  $\text{Head} > \text{Comp}$  [i.e.  $\{<\alpha, \beta>, <\beta, \alpha>\} \rightarrow \{<\beta, \alpha>\}$ ]

In order to argue for the superiority of his proposed account over previous proposals, special emphasis is put by Richards also on the empirical outcome of (48) for linearization patterns. In particular, according to the author, «the linear order-preservation effect known as Holmberg’s Generalization (HG) that constrains Germanic Object Shift is directly entailed by the ‘head-initial’ (VO) setting [48-a] of this parameter» (Richards 2008, p. 150). Taking Object Shift (OS) to be «the short leftward displacement of weak (destressed) objects in Germanic» and HG «the constraint on this operation such that the

shifted object cannot cross an in-situ (nonfinite) lexical verb», the issue which Richards alludes to concerns the fact that HG holds only for VO languages such as Icelandic (Richards 2008, p. 150). For example, while in this latter languages OS cannot apply unless V2-raising has occurred (as shown in (49)), in OV systems such as German «the equivalent short object movement in OV Germanic ('scrambling') may occur irrespective of the finiteness of the main verb» (as shown in (50)) (Richards 2008, pp. 150-151):

- (49) a. Nemandinn las (bókina) ekki (bókina)  
 The-student read (the-book) not (the-book)  
 "The student read didn't read the book."  
 b. Nemandinn hefur (\*bókina) ekki lesid (bókina)  
 The-student has (the-book) not read (the-book)  
 "The student hasn't read the book."
- (50) a. Der Student las (das Buch) nicht (das Buch)  
 The-student read (the-book) not (the-book)  
 "The student didn't read the book."  
 b. Der Student hat (das Buch) nicht (das Buch) gelesen  
 The-student has (the-book) not (the-book) read  
 "The student hasn't read the book."

According to Richards (2008), the fact that HG only applies to VO languages can be straightforwardly accounted for by assuming it to be a verb-object order preservation constraint deriving from (48a). Considering the syntactic configuration created by OS over V in a VO language (Richards 2008, p. 151):



Richards's intuition is that, since in VO systems the parametric setting (48a) instructs PF to systematically ignore all c-command relations  $O > V$ , the precedence instruction corresponding to the  $O > V$  chain established by movement of the object across the lexical verb would be deleted before Spellout. Hence, the only way for the shifted object to be licensed is that there be V-to-T raising, which restores the c-command relation – and thus the non-vacuous precedence instruction –  $V > O$ . In this way, according to Richards, «HG is derived for exactly that subset of languages in which it holds (namely, those set to [48-a], i.e. VO languages)», while in OV languages scrambling may apply regardless of whether the main verb is raised or not (Richards 2008, p. 151).

One problem, however, with Richards's order-preservation account of HG is that, as stated by the author himself, «order preservation is by far the exception rather than the rule» (Richards 2008, p. 151), as it can be seen not only when considering verb movement, as in (50a), but also typical instances of A/A-bar movement such as passivization, topicalization, and wh-movement, as in (52) (Richards 2008, p. 152):

- (52) a. A man arrived (*a man*)  
       b. John was rescued (*John*)  
       c. John, I like (*John*)  
       d. Which book did you read (*which book*)

Richards's proposed solution to this problem is based on two independently-motivated assumptions: first, that «all varieties of  $v$  are phase heads – i.e. both transitive  $v^*$  and passive/unaccusative 'defective'  $v$  ( $v_{\text{def}}$ )»; and second, that «OS/Scrambling targets spec-vP», as in (51) (*ibidem*). By comparing order-preserving object movement such as OS/scrambling in (49) and (50b) with order-disrupting object movement such as (52), under this perspective it is clear that, while the former does not cross a phase boundary, the latter targets a position higher than Spec-vP. These two kinds of movement can thus be characterized by means of the following generalization (*ibidem*):

- (53) a. Order-preserving movement is phase-internal.  
       b. Order-disrupting movement is cross-phasal.

In these terms, Richards continues, «all that is required in order to derive (53) is for (48) to operate on a phase by-phase basis» (Richards 2008, p. 153). This means that, in order to account for both External and Internal Merge, Parametrized desymmetrization

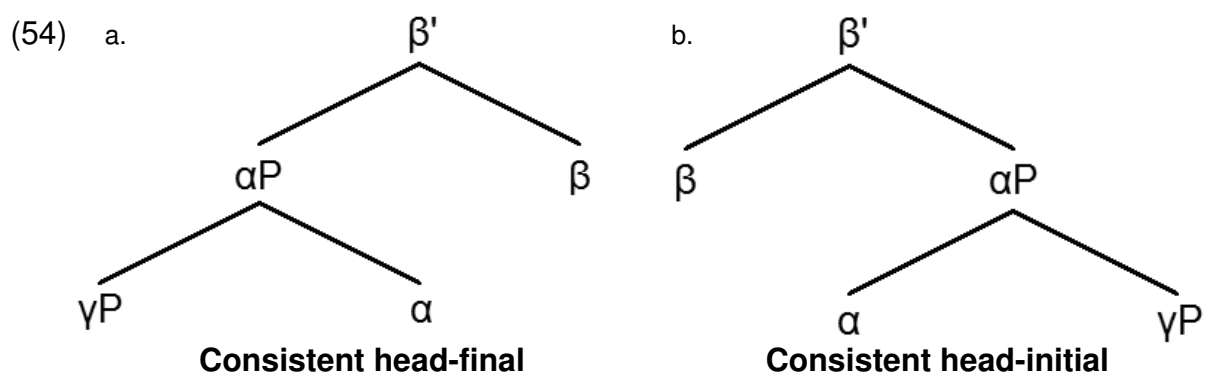
must hold up to the phase level. According to Richards, this is not an unlikely scenario, as it precisely conforms to Chomsky's (2001b) view of Spell-out as applying at every phase level to minimize the computational burden on working memory.

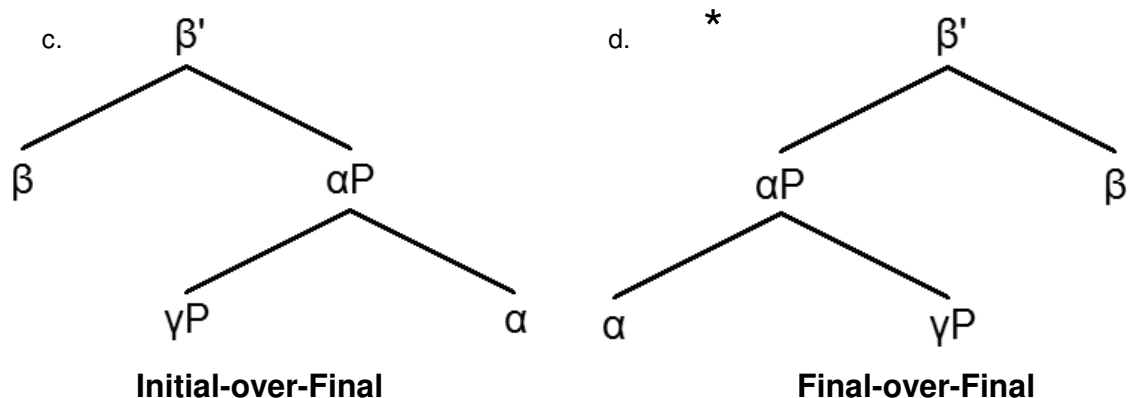
In conclusion, Richards (2004, 2008) represents a notable attempt to actualize the head-complement parameter, essentially by parametrizing a weaker, non-asymmetrical version of Kayne's (1994) LCA. As a result, Richards not only shifts the locus of this traditional parameter from UG to the interface conditions holding between narrow syntax and the phonological system but, by basing his account on a symmetrical view of phrase structure, also emphasizes the minimalist idea of language faculty in the narrow sense as not perfectly designed with respect to the requirements imposed by the AP interface.

### 5.3.2 – Head directionality, narrow syntax, and the Final-over-Final Constraint

If on the one hand Richards's (2008) reformulation of the head-complement parameter conforms to the dominant minimalist view according to which linear order is established outside narrow syntax, this is not the only approach which is currently being adopted in Generative Grammar. In this respect, one notable example of a parametric account resorting to the idea that linearization is a matter of narrow syntax is the one proposed in Biberauer & Roberts (2015), which in turn builds on Biberauer, Holmberg & Roberts's (BHR 2009, 2014) analysis of word order asymmetries.

Differently from Richards (2004, 2008), who proposes to adopt a revised version of Kayne's LCA mainly on theoretical grounds, the hypothesis originally put forth in BHR (2009) and developed as a full-fledged syntactic parameter in Biberauer & Roberts (2015) is based on empirical observations concerning what BHR refer to as «a skewing in the word-order patterns attested in the world's languages» (BHR 2009, p. 78). This asymmetry is represented in (54) (*ibidem*):





According to the authors, the existence of the skewing in the disharmonic domain shown above is supported by extensive cross-linguistic data; in these terms, they note, no language seems to adopt the disharmonic configuration (54d) (cf. BHR 2009, p. 78). Before introducing some of the pieces of empirical evidence provided in this respect by BHR, however, a brief look is given to the generalization whereby the authors propose to account for this word order asymmetry. This generalization takes the form of the condition given in (55), which is referred to as the Final-over-Final Constraint (FOFC). The formulation proposed here is taken from Biberauer & Roberts (2015):

(55) The Final-over-Final Constraint:

A head-final phrase  $\beta P$  cannot dominate a head-initial phrase  $\alpha P$ , where  $\alpha$  and  $\beta$  are heads in the same Extended Projection. (Biberauer & Roberts 2015, p. 305)

The direct effect of (55) is that the following configuration is invariably ruled out:

(56) \* $[\beta P \dots [\alpha P \dots \alpha \gamma P] \beta \dots]$

As pointed out by the authors, not only does this generalization overarch on all language types, but also on all categories. Moreover, it should be borne in mind that, although having clearly observable effects on linear order, «FOFC is a universal constraint on phrase structure configurations, not statable in purely linear terms» (BHR 2014, p. 170).

As far as relevant empirical evidence is concerned, the first set of data presented by BHR in favor of FOFC concerns «the apparent crosslinguistic absence of VOAux orderings» (BHR 2009, p. 78). In this respect, as noted by the authors, examples of the orders OV-Aux and Aux-OV can easily be found in spoken Afrikaans (as shown in (57a-b)), while the head-initial order Aux-VO is allowed in English-influenced Kaaps (as shown

in (57c)). Crucially, however, «no variety of Afrikaans [...] allows [57d], the FOFC-violating order» (BHR 2009, p. 79):

- (57) a. ...dat sy [<sub>DP</sub> 'n brief] *geskryf* **het** [OV-Aux]  
           that she a letter      written has  
           “...that she has written a letter.”  
       b. ...dat sy **het** [<sub>DP</sub> 'n brief] *geskryf* [Aux-OV]  
       c. ...dat sy **het** *geskryf* [<sub>DP</sub> 'n brief] [Aux-VO]  
       d. \*...dat sy *geskryf* [<sub>DP</sub> 'n brief] **het** [VO-Aux]

Regarding the fact that the FOFC-violating word order VO-Aux can still be observed in some languages, according to BHR these counterexamples are only apparent. In fact, what these cases actually show is that «VOAux structures are exclusively permitted in languages featuring non-inflecting auxiliary elements, commonly designated *particles*», as shown in the following examples from Bwe Karen (*ibidem*):

- (58) a. yə- ca dəyo lə (Dryer 2009)  
           1SG-see picture ASP  
           “I am looking at a picture.”  
       b. ceçə mi jə-khoʻ phi má nɔ (\*jə-kho)  
           3- say COMP 3- FUT take what  
           “What did he say that he would take?”

Examples such as (58) seem, therefore, to be explainable by formally distinguishing inflecting auxiliaries from non-inflecting auxiliary particles, an idea which seems to be confirmed by the observation that «non-inflecting particles expressing tense-aspectual (auxiliary) information necessarily occupy a very different position to that in which inflecting auxiliaries obligatorily surface» (*ibidem*).

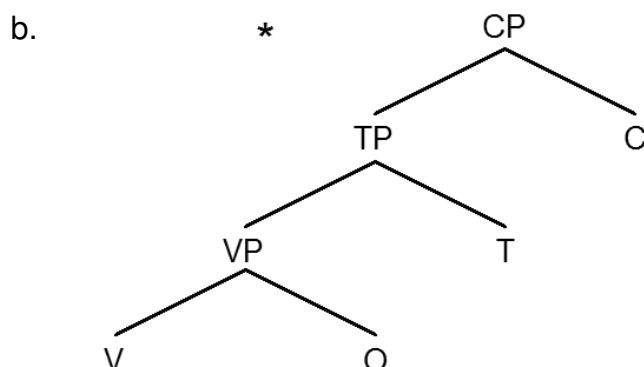
The second source of empirical evidence mentioned by BHR concerns «the absence of VO languages with final complementizers» (BHR 2009, p. 78; the original text refers to initial complementizers, but final complementizers seems to be correct in this context). In particular, the authors focus their attention on final adverbial subordinators, whose sentence-final occurrence in VO languages appears to be far more restricted than the occurrence of their sentence-initial counterparts in OV languages:



According to the on-line *World Atlas of Language Structures* (WALS; Haspelmath et al. 2008), 54 clearly OV languages (out of a sample of 64) feature an initial adverbial subordinator (“because”), and this excludes familiar OV languages like German and Dutch, which, on account of their matrix V2 property, are listed as languages with “no dominant order”; only 2 VO languages are said to feature final adverbial subordinators, but see Newton (2008) for a critique of the descriptions underlying this classification. (BHR 2009, p. 80)

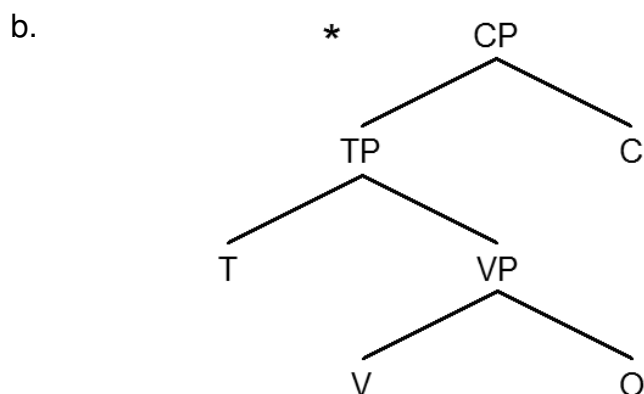
Assuming with Grimshaw’s (2000) that «the nominal system and the verbal system form (extended) projections, which include both the projection of their lexical heads and the functional shell which surrounds the lexical projection» and which, in these terms, T and C are part of the extended projection of V, the ban on final complementizers imposed by VO languages can be accounted for in two ways (Grimshaw 2000, p. 116). On the one hand, FOFC can be violated by a head-final TP immediately dominating a head-initial VP, as shown in (59) (BHR 2009, p. 80):

- (59) a.  $*[_{CP} [_{TP} [_{VP} V O] T] C]$  -- violates FOFC for  $\alpha = V$  and  $\beta = T$



On the other hand, FOFC could also be violated by a head-final TP being immediately dominated by a head-initial CP, as shown in (60) (*ibidem*):

- (60) a.  $*[_{CP} [_{TP} T [_{VP} V O]] C]$  -- violates FOFC for  $\alpha = T$  and  $\beta = C$



The last type of evidence provided by BHR in support of the universal nature of FOFC concerns «the diachronic pathways which languages appear to follow during the process of word-order change» (BHR 2009, pp. 78-79). Starting from the assumption that FOFC represents a universal constraint, BHR prediction is that word order change cannot possibly generate a FOFC-violating system, but rather will strictly proceed along one of the two following diachronic pathways (BHR 2009, p. 81):

- (61)  $[[[O\ V]\ I]\ C] \rightarrow [C\ [[O\ V]\ I]] \rightarrow [C\ [I\ [O\ V]]] \rightarrow [C\ [I\ [V\ O]]]$   
 (62)  $[C\ [I\ [V\ O]]] \rightarrow [C\ [I\ [O\ V]]] \rightarrow [C\ [[O\ V]\ I]] \rightarrow [[[O\ V]\ I]\ C]$

According to the authors, failure to comply with either (61) or (62) results in a violation of FOFC, and thus is not a viable diachronic option. As noted by BHR, literature on different language families reports that diachronic change indeed follows the pathways outlined in (61) and (62), thus providing additional support to the FOFC:

Biberauer, Newton & Sheehan (2009a,b [=2009,2010, A.R.]) present case studies from the history of Germanic and Ethiopian-Semitic, showing that directionality changes in languages belonging to these families conform to the expected FOFC-determined pathways. Further, it appears that Niger-Congo languages that have undergone varying amounts of initial to final change have likewise done so in the predicted “bottom-up” fashion (cf. Nikitina 2008 for recent discussion); similarly, the Sami languages within Finno-Ugric, appear to have undergone the reverse change “top-down”. (*ibidem*)

At this point, BHR introduce a further aspect of their theory which, crucially, constitutes a major determinant of the nature of the parametric proposal which would be formulated in Biberauer & Roberts (2015). This aspect concerns the fact that, since FOFC is supposed to hold for two phrases which belong to the same extended projection, the very skewing characterizing the relevant data reviewed so far is predicted not to arise when considering a potentially FOFC-violating relation intercurring between two phrases belonging to different extended projections. For example, the German sentence (63) shows how a head-final VP can dominate a head-initial DP without triggering a violation of FOFC (*ibidem*):

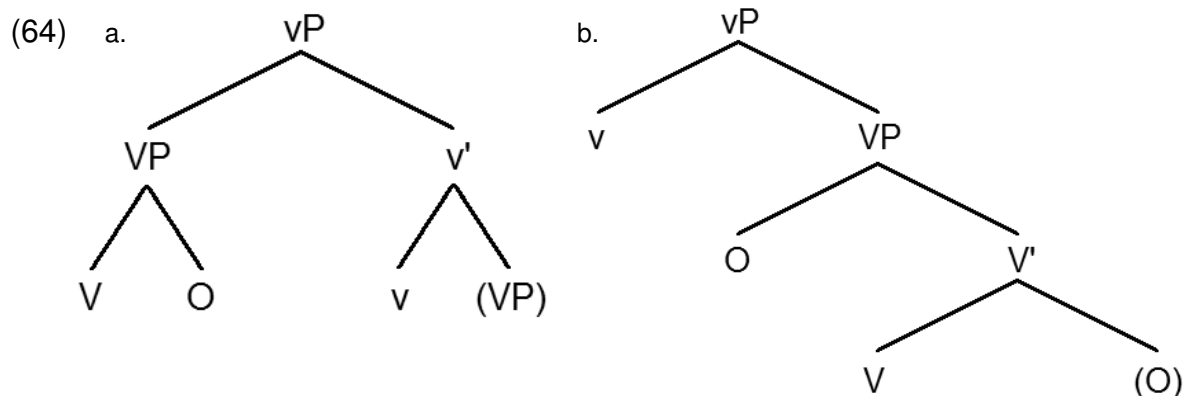
- (63) ...dass sie gerne [<sub>DP</sub> ein Täschen Kaffee] trinkt  
           that she gladly a cup-DIM coffee drinks  
           “...that she enjoys drinking a cup of coffee.”

This contrast is very important because, according to BHR, the non-total nature of the FOFC «make[s] it unclear how a Head Parameter (HP)-based account would be able to rule out the problematic orders without stipulation» (BHR 2009, p. 82):

In the minimalist context, this is significant: as noted in the introduction, the view that linearization information, like phonological specifications more generally, has no place in NS and should therefore only be imposed at PF, possibly via a PF parameter, is widespread (cf. Biberauer 2008a [=2008, A.R.] and Richards 2009 [=2008, A.R.] for recent overview discussion and references). It is our contention, however, that the empirical facts mentioned above and the apparent universality of [55] fatally undermine the validity of this assumption. (*ibidem*)

Discarding, therefore, an account based on a PF parameter in the spirit of Richards (2004, 2008), BHR's conclusion is that, «contrary to what is arguably the dominant view in minimalist theorizing today, [...] linearization information must in fact be specified Narrow Syntax-internally» (BHR 2009, p. 77).

Although the theoretical model proposed in this paper is slightly outdated compared to the one formulated in Biberauer & Roberts (2015), BHR (2014) accounts straightforwardly for the attested skewing in linear order pointed out in BHR (2009). Starting from the assumption that «disharmonic orders result when some complements, and/or elements contained in those complements, undergo movement and others do not», BHR propose FOFC to be a ban on impossible linearization strategies which in turn arise «when a superordinate head triggers movement of its complement, but inside that complement the head does not trigger movement of *its* complement». (BHR 2014, p. 208). A comparison between a FOFC-violating configuration (64a) and a FOFC-compliant one (64b) is given below:



Assuming that *v* can contain an auxiliary, then (64a) could be linearized as VO-Aux, that is, precisely the impossible word ordering discussed in BHR (2009). As pointed out by BHR, the ordering given by this structure is the result of the fact that, while *v* triggers roll-up movement of VP to Spec-vP, *V* does not do the same with its complement *O*, which therefore stays *in situ*. On the other hand, in (64b) *v* does not trigger movement of its complement but *V* does, hence the surface order Aux-OV. As these examples show, a FOFC-violating construction arises when, given an extended projection which is neither consistently head-initial or consistently head-final, all the phrases below the highest head-final phrase are not consistently head final (cf. BHR 2014, p. 209).

At this point, BHR's solution to this problem is that «FOFC follows from the interaction of the antisymmetric nature of UG (i.e., the fact that head-final order requires  $\wedge$ )» – a symbol whose meaning will be clarified immediately below – «combined with the highly local nature of c-selection, and Relativised Minimality» (Biberauer & Roberts 2015, p. 305). BHR's central assumption is that «the head in a head-final phrase must have a feature triggering movement», which they represent with the symbol  $\wedge$ , «of its complement, which a head-initial counterpart does not have» (BHR 2014, p. 172). Essentially, what the authors propose in this respect is that FOFC is a by-product of the syntactic process whereby the movement-triggering feature  $\wedge$  spreads through the functional heads of a given extended projection. Some of the properties of  $\wedge$  are listed below:

Unlike  $\phi$ -features, which are arguably best seen as attribute-value pairs,  $[\wedge]$  is privative, it has no internal structure, it cannot be valued or in any obvious way “checked off,” and [...] it has no semantic or morphophonological effects. (BHR 2014, pp. 209-210)

As well as displaying such peculiar properties, the movement-triggering feature  $\wedge$  postulated by BHR is inherited via c-selection (cf. BHR 2014, p. 211). According to this proposal,  $\wedge$  spreads through the extended projection together with the categorial feature of the lowest lexical category. In the case at hand, for example, when *v* c-selects  $[+V]$  and merges with *V(P)*, although «*v* is not inherently valued  $[+V]$ , [...] this feature “spreads” to *v* from its sister VP» and so on, thus defining the extended projection itself (*ibidem*). Most importantly, the inheritance of  $\wedge$  via c-selection is, according to BHR, strictly dependent on the inheritance of a categorial feature. What this means is that, while on the one hand  $[+V]$  can percolate upwards without  $\wedge$ , on the other hand  $\wedge$  cannot be transferred on its own:

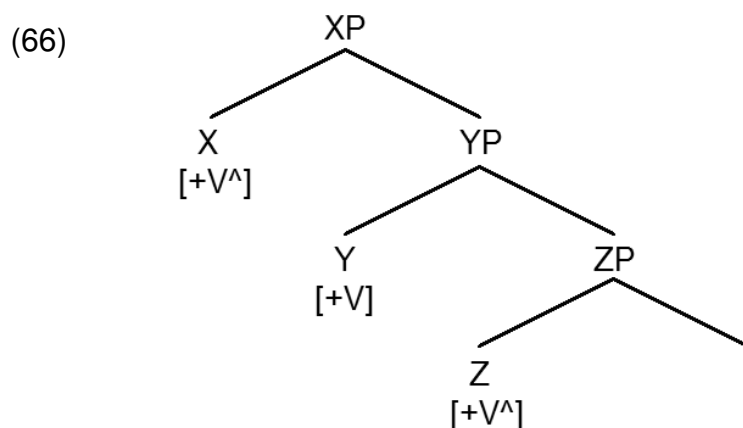
What is most important for our purposes concerns the interaction of  $\wedge$  with selection. If a given head can select  $[+V]$  and inherit  $[+V]$ , exactly the same applies in a system with  $[+V^\wedge]$ . In this situation, a higher head may select  $[+V]$  and inherit  $[+V]$  without  $\wedge$ , but, *crucially, no head can inherit  $\wedge$  without inheriting  $[+V]$* . The assumption behind this is that  $\wedge$  cannot be selected alone, since it is not a categorial feature. (BHR 2014, p. 211)

In addition to the fact that «parametric variation in word order can then be encoded in terms of the highest head in the extended projection that selects  $[\pm V^\wedge]$ », another desirable consequence of BHR's model is that, as anticipated above, FOFC can be described as following from the locality of c-selection (*ibidem*). In this respect, this specific requirement is argued to be an effect of Rizzi's (2001) Relativized Minimality, which BHR state as follows (BHR 2014, p. 212):

(65) Relativized Minimality (adapted from Rizzi 2001)

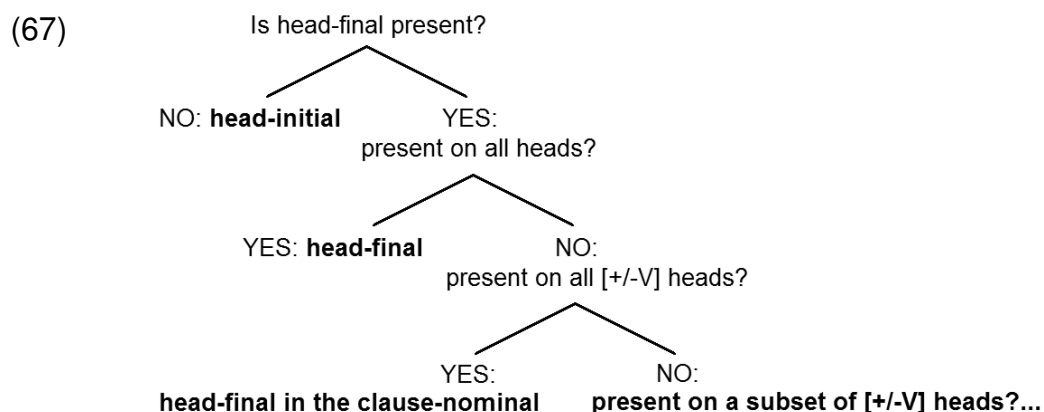
In a configuration  $X \dots Y \dots Z$ , where  $X$  asymmetrically c-commands  $Z$ , no syntactic relation  $R$  can hold between  $X$  and  $Z$  if  $Y$  asymmetrically c-commands  $Z$  but does not c-command  $X$ , and  $R$  potentially holds between  $X$  and  $Y$ .

The effect of (65) on a prototypical FOFC-violating configuration is clearly illustrated in (66) (*ibidem*):



In (66),  $Y$  intervenes between  $X$  and  $Z$  and, therefore,  $X$  cannot c-select  $Z$ . Hence, since  $\wedge$  cannot spread from  $Z$  to  $X$  without being also specified on  $Y$ , the FOFC-violating structure (64a) cannot be derived (cf. *ibidem*).

Although BHR (2014) do not embark upon a parametric account of word order, such task is undertaken by Biberauer & Roberts (2015). In this paper, the format of head-complement parameter is argued to adhere to the hierarchical model put forth in Roberts & Holmberg (2010) (Biberauer & Roberts 2015, p. 301):



Following BHR (2014), the feature whose distribution is held responsible for the emergence of the hierarchy in (67) is the diacritic  $\wedge$  which, «when associated with the categorial feature of a head, triggers movement of the complement of that head to its specifier» (Biberauer & Roberts 2015, p. 301). Although not consisting in an unvalued morphosyntactic feature of the kind which was proposed to account for null subject and verb movement phenomena, the movement-triggering feature  $\wedge$  undergoes the same process of input generalization. In these terms, at the top of the hierarchy there is the first macroparametric option with  $\wedge$  being entirely absent. This setting corresponds to harmonically head-initial languages, which are therefore characterized as maximally-unmarked systems. Then, the next least marked (macroparametric) option is selected when « $\wedge$  is assumed to apply throughout the system so that a fully head-final system emerges» (Biberauer & Roberts 2015, p. 304).

Proceeding further down the hierarchy, the next parametric choice imposes selection on which kind of extended projection the feature  $\wedge$  must be uniformly associated with, the two versions available differing in their categorial specification, being it either [+V] or [-V]. In the case at hand, the option considered by Biberauer & Roberts (2015) is the extended verbal projection, which they assume to concern the phase heads C and v. At this point, the authors argue, «if the PLD [primary linguistic data] is such that generalized order (either head-final or head-initial) inside the verbal Extended Projection cannot be maintained (i.e., if it is such that some strings must be parsed as head-final and others as head-initial), then the next categorial distinction is posited», with C and v being characterized as [+V,+C] and [+V] categories respectively (*ibidem*).

Biberauer & Roberts's word order hierarchy contemplates further marked options, which are implemented by positing an additional distinction between the core functional categories C, T, v before embarking on micro- and nanoparametric variation. While not delving deeper into these matters, the authors nonetheless make their position clear with

respect to both their view of linearization and their idea of parametric variation. Differently from Richards (2004, 2008), not only do Biberauer & Roberts (2015) support the idea that linearization is a product of narrow syntax, but they also adopt for head-directionality the same emergent model which has been applied, although with some minor differences, to the majority of the once-GB parameters which are now still accepted in Minimalism.

### 5.3.3 – *An experimental perspective on linearization*

Before drawing any final conclusions from our historical review of the head-complement parameter, some results from a recent experimental study will be evaluated to better understand the empirical implications of the theories of linearization which have been considered in this chapter. In this regard, Donati & Branchini (2013) seems to be a particularly fitting choice in this respect as its aim and scope precisely encompass the perspectives of the three main schools of thought on the parameterization of head directionality seen so far: the narrow syntactic view (Travis (1984), Biberauer & Roberts (2015)), the “Kaynean” view (Kayne (1994)), and the “Chomskyan” view (Richards (2004, 2008)).

Donati & Branchini (2013), whose title is *Challenging linearization: simultaneous mixing in the production of bimodal bilinguals*, analyzes data from six Italian CODAs (Children of Deaf Adults) between the ages of 6 and 8 years who are native users of Italian Sign Language (LIS) and Italian. In their analysis, Donati & Branchini focus in particular on these children's use of code-blended utterances, that is, speech-sign productions in which bimodals, by combining oral/auditory and gestural/visual information, «do not need to stop talking in order to sign, or vice-versa, but may speak and sign simultaneously» (Donati & Branchini 2013, p. 97).

Before delving into the focus of their work, Donati & Branchini briefly outline some basic syntactic properties of LIS, along with some code-mixing phenomena which characterize the speech of their experimental bimodal bilingual subjects, in order to better contextualize their proposal. This is done primarily by evaluating a set of examples from LIS, whose signs are glossed in their citation form with capitalized English words. Most interestingly for their paper's purposes, although Italian and LIS share some grammatical properties such as pro-drop and rather free constituent dislocation for information-related purposes, «while Italian is a harmonic head-initial language, LIS is a harmonic head-final language with negation, modals, and even wh-elements following the verb» (Donati &

Branchini 2013, p. 99). LIS's basic word order is SOV, as shown in (68) (Donati & Branchini 2013, p. 101):

- (68) MARY HOUSE BUY  
 "Mary buys a house."

As noted by the authors, «LIS displays a number of clausal functional categories all occupying a post verbal position», that is, «the aspectual marker DONE, marking that the action expressed by the verb has been completed [69a], modals [69b], negation [69c], manner adverbs [69d] and wh-phrases [69e]» (Donati & Branchini 2013, pp. 101-102):

- (69) a. MARY HOUSE BUY DONE  
 "Mary has bought a/the house."  
 b. MARY HOUSE BUY CAN  
 "Mary can buy a/the house."  
 c. MARY HOUSE BUY NOT  
 "Mary doesn't buy a/the house."  
 d. MARY ARRIVE LATE  
 "Mary arrives late."  
 e. MARIA BUY WHAT  
 "What does Mary buy?"

According to the definition given by Gumperz (1982), code mixing represents «the juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems» (Gumperz 1982, p. 59). An example is provided in (70) (Donati & Branchini 2013, p. 103):

- (70) It:                      e      poi    l'      ha              preso  
                                  and   then   it      have.3SG   take.PTCP  
 LIS: CUT-HEART                                      TAKE-HEART  
 "(He) has cut the heart and has taken it."

While «speech-sign bilinguals rarely code-switch», a rather common type of production for Donati & Branchini's bimodal subject consists in simultaneously mix bits of



the two languages – a peculiar phenomenon which is commonly referred to as code-blending (Emmorey *et al.* 2005, p. 665). The first subtype of code blending focused on by Donati & Branchini is dominant blending. In dominant blending, one language provides an autonomous and complete utterance while the other offers a marginal support to the global utterance (Donati & Branchini 2013, p. 104):

- (71) It:    La    strega        dà            la    mela    a    Biancaneve  
           The   witch        give.3SG    the   apple   to   Snow White  
      LIS:                            CL-GIVE  
           “The witch gives the apple to Snow White.”

The second type of blending is represented by *independent blending*, which «involves the simultaneous production of two independent and apparently autonomous monolingual utterances» (Donati & Branchini 2013, p. 105). As the authors note, some slight differences between the two utterances can be observed. This difference may in turn lie in the fact that, for example, «the utterance may be semantically richer in Italian, as in [72] or in LIS, as in [73]» (*ibidem*):

- (72) It:    Lavora                    a    Rimini  
           work.PRES.3SG    in    Rimini  
      LIS:   WORK                    LOC  
           “He works there, in Rimini.”
- (73) It:    I                    sette    nani        sono                    saliti  
           The.PL        seven    dwarf.PL    be.PRES.3PL        climb-PTCP  
      LIS:   SEVEN DWARVES                                    CLIMB ON-SHOULDERS  
           “The seven dwarves have climbed on the shoulders.”

Given the word order difference mentioned above between Italian and LIS, one question raised at this point concerns how the two independently-blended utterances are linearized in the global utterance. As far as their corpus data are concerned, the authors list three possibilities, the first of which is what they refer to as *congruent lexicalization*, that is, the procedure whereby «the two blended strings happen to display the same word order in the two languages according to their specific grammars», avoiding any

discrepancies in their respective linearization (Donati & Branchini 2013, p. 107). An example is given below (*ibidem*):

- (74) It: Lei sa tutto  
           she know.3SG everything  
       LIS: IX KNOW ALL  
           “She knows everything.”

Crucially, this procedure is possible for (74) as «universal quantifiers are a documented exception to the strong OV order of LIS [...], and systematically follow the verb» (*ibidem*). This specific pattern is thus very infrequent in the experimental data.

The second strategy employed by Donati & Branchini’s CODAs is to resort to syntactic calques. In this case, both the two blended strings strictly conform to the word order imposed by one of the two languages. In examples (75) and (76), the global linear order is the one imposed by Italian and LIS respectively (Donati & Branchini 2013, p. 108):

- (75) It: Una bambina va allo zoo  
           A girl go.3SG to.the zoo  
       LIS: GIRL GO ZOO  
           “The girl goes to the zoo.”
- (76) It: Il Papà la mamma la sorella mangiato finito  
           The Father the mother the sister eat.PTCP finish.PTCP  
       LIS: FATHER MOTHER SISTER EAT DONE  
           “The father, the mother and the sister have done eating.”

The third and final solution for independent blending is «when each of the two utterances follows the typical word order prescribed by its respective language» – an option which produces bimodal utterances with contradictory Italian and LIS word orders (*ibidem*). This pattern is illustrated in (77-78) (Donati & Branchini 2013, p. 109):

- (77) It: Eh? Non ho capito  
           Uh? NEG have.1SG understand.PTCP  
       LIS: I UNDERSTAND NOT  
           “I haven’t understood.”

- (78) It: Chi ha chiamato?  
           who have.3SG call.PTCP  
       LIS: CALL WHO  
           “Who has called?”

In (77), the difference between the Italian and LIS concerns the placement of negation, which in the former language precedes the verb while in the latter language follows it. In (78), on the other hand, the difference concerns the *wh*-element, which in Italian questions is clause-initial while in LIS is clause-final.

Returning to the typology of code-blending encountered in their data, after dominant blending and independent blending Donati & Branchini focus their attention on what they call *blended blendings*. A blended blending is a bimodal mixed utterance which «is complete and meaningful only if the fragments distributed in the two channels are put together in a unique, blended utterance» (Donati & Branchini 2013, p. 110). This typology of blending is exemplified in (79) and (80) (*ibidem*):

- (79) It: Parla con Biancaneve  
           talk.PRES.3SG with Snow White  
       LIS: TALK HUNTER  
           “The hunter talks to Snow White.”
- (80) It: dalla regina cattiva  
           to.the queen wicked  
       LIS: GO WICKED  
           “(He) goes to the wicked queen.”

In (79), while the verb is produced in both utterances, Italian provides the indirect object while LIS provides the subject. In (80), on the other hand, Italian provides the locative argument and LIS the verb. In this respect, the most striking aspect of blended utterances such as (79-80) is that, as shown in the glosses, the fact that there are pairs of constituents which are realized in simultaneity means that, contrary to what happens in unimodal utterances, such constituents have not been linearized (cf. Donati & Branchini 2013, p. 111).

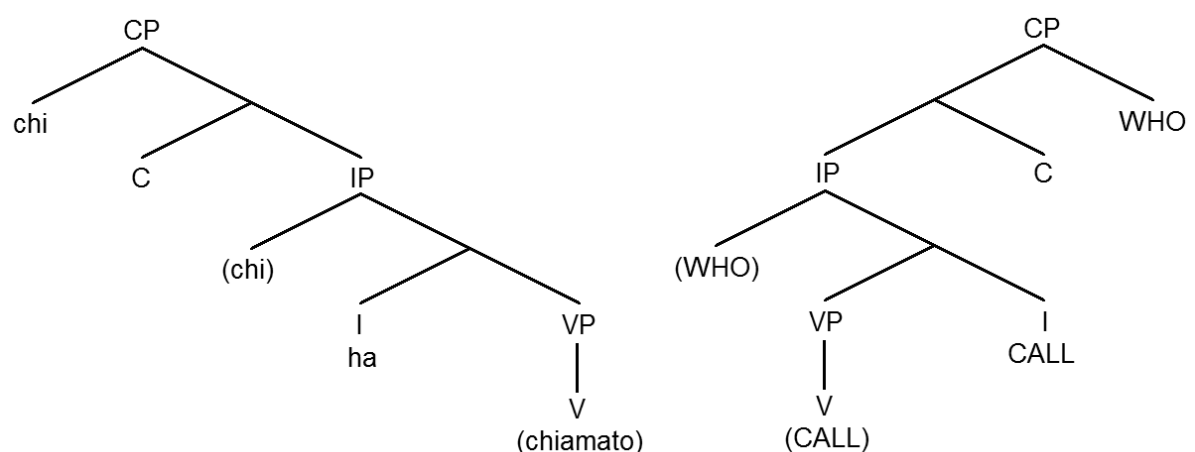
At this point, Donati & Branchini confront the mayor theories of linearization with a specific set of their data, that is, the observed cases of independent blending where the

sign and the speech follow two contradictory word orders, as in (77-78). The first hypothesis to be evaluated in this respect is the one corresponding to the head-parameter which was first systematized by Travis (1984). Starting with Travis's theory of linearization, given that head-directionality was assumed to be entirely determined at phrase structure level within the GB-model, one implication of Travis's hypothesis is that two strings displaying two different linear orders necessarily derive from two distinct phrase structures.

Differently from Travis (1984), the revised version of the head-complement parameter put forth by Biberauer & Roberts (2015) allows to derive not only harmonic word orders but also disharmonic ones by relativizing head directionality, which is now category-specific rather than language-specific. However, although representing significant improvement over its pre-Minimalist predecessor, Biberauer & Roberts's proposal can still be assimilated to Travis's account as, like the latter, it still implies linear ordering to be established within the syntactic module, and hence the impossibility of having more linearization patterns from the same syntactic structure.

Regarding the specific pieces of production considered in the paper at issue, a blend displaying two contrasting word orders such as (78) could be accounted for, under these narrow-syntactic views, only by postulating «the simultaneous base generation of two structures which are the mirror images of each other», as shown in (81) (Donati & Branchini 2013, p. 118):

(81)

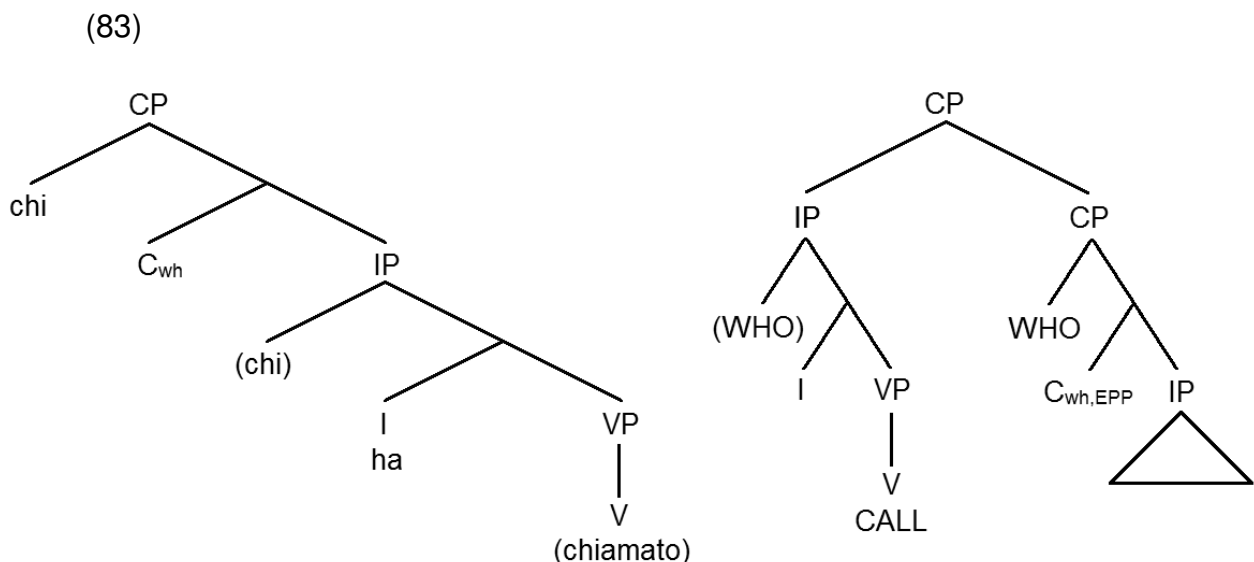


The second theory of linearization referred to by Donati & Branchini is the one represented by Kayne's (1994) LCA and other similar approaches, which hypothesize the existence of a universal, non-parametrized algorithm mapping structural relations holding at the narrow syntactic level to linear precedence. Specifically for the purposes of their

analysis, the most important aspect of this model is that, while the LCA corresponds to a linearization algorithm operating at PF, it still has a syntactic effect as «structures that are not linearizable by such an algorithm are simply not generable» (Donati & Branchini 2013, p. 115). Starting from the assumption that, in these terms, Specifier-Head-Complement/SVO is the only basic order allowed, «if two language strings superficially display two different linear orders, they necessarily correspond to two different syntactic structures, one involving movement operations not included in the other» (*ibidem*). This is exemplified in (82), in which the harmonically head-final word order (82b) is, contrary to the harmonically head-initial string (82a), transformationally derived as in (82c) (*ibidem*):

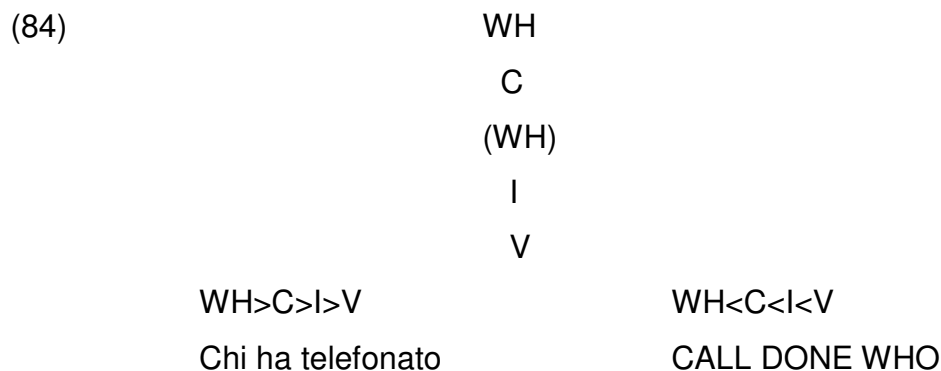
- (82) a. C I V D NP  
 b. NP D V I C  
 c. [CP [IP[VP[DP NP D t<sub>NP</sub>] V t<sub>DP</sub>] I t<sub>VP</sub>] C t<sub>IP</sub>]

Consequently, the only way for such an approach to account for (78) would be to assume that «such data are generated through the parallel simultaneous computation of two syntactic derivations starting from two identical base structures», with each channel giving as its output an independently-linearized syntactic structure (Donati & Branchini 2013, p. 118). This is shown in (81) (Donati & Branchini 2013, p. 119):



Finally, the third hypothesis address by the authors is the one which Donati & Branchini conventionally refer to as “Chomskyan”, a label which subsumes all those analyses according to which linearization is regarded as an process taking place at the

sensori-motor interface/PF and not affecting either syntactic derivations nor the latter's semantics. Differently from both the narrow syntactic view and the "Kaynean" view briefly outlined above, this hypothesis implies that, as noted by the authors, «two language strings displaying two divergent word orders might correspond to the same abstract syntactic structure, linearized by different settings of some linearization parameter(s)» (Donati & Branchini 2013, p. 115). In these terms, independently-blended utterances such as (78) can be accounted for by deriving each string from the very same syntactic structure, which is linearized twice according to each of the two distinct linearization algorithms pertaining to each of the two interfaces involved in bimodal communication. A schematic representation of this procedure is given in (84) (Donati & Branchini 2013, p. 120):



Crucially, what emerges from this evaluation is that, while the first two hypotheses are based on the generation of two parallel syntactic derivations in order to account for the observed cases of two word order blending, the latter hypothesis can yield the same result by associating two distinct PF-representations with one and the same syntactic structure. As pointed out by the authors, these two scenarios require us to examine two alternative hypotheses: on the one hand, that «bimodals possess an ability not obviously connected to the availability of two PF channels, namely that of being able to compute two syntactic structures simultaneously»; on the other hand, that the realization of bimodal utterances follows directly from «the availability in bimodality of two channels and hence two PF spaces», with linear order variation being therefore due to a difference in the linearization algorithms imposed by the oral/auditory and the gestural/visual channels respectively (*ibidem*).

In order to check these different predictions and settle things once and for all, Donati & Branchini take into account blended blending utterances, whose simplest example occurring in their corpus is represented by (85) (Donati & Branchini 2013, p. 123):

(85) It: io!  
I  
LIS: WIN  
“I won!”

What blended utterances such as (85) and (79-80) show is that the constituents parallelly distributed across the two channels by Italian and LIS form a coherent global utterance even if they are realized simultaneously. Linearization, in other words, ultimately seems not to be required either syntactically nor semantically. In fact, as noted by the authors:

[...] what the examples discussed above [...] show clearly is that constituents such as subject and object, object and indirect object, verb and locative, etc., do not need to be linearized in order to be integrated into a blended syntactic structure and get compositionally interpreted. (*ibidem*)

Therefore, according to Donati & Branchini, the third hypothesis is the only one which can give a viable account of the existence of blended blendings: assuming linearization to be purely phonological and, as such, completely autonomous from syntax and semantics, this specific kind of blending can be seen as an effect of the exceptional availability of two SM interface channels, which partially suspends the linearization requirement which is typically forced by the availability of a single articulatory channel:

If linearization is considered to be phonological, something that happens to syntactic structures in order to adapt them to the linear articulatory constraints connected to the sensory-motor interface, the existence of this type of blending is straightforwardly accounted for: due to the exceptional circumstances of the availability of two linear channels instead of one, these usual articulatory constraints can be partially suspended and linearization partially overridden. (Donati & Branchini 2013, pp. 123-124)

Conversely, «if linearization is considered to play a more central role in relation to syntax», as argued for by Travis (1984) and Biberauer & Roberts (2015), «these data

become extremely problematic» (Donati & Branchini 2013, p. 124). In these terms, given the fact that they occupy distinct structural positions, two syntactic constituents would be never allowed to occupy the same position in linear order. As the authors suggest, these data pose a most critical problem also for the Kaynean stand, as «if linearization is a filter on syntactic structures you predict simultaneity to be impossible» (*ibidem*). From this perspective, any instance of simultaneity occurring in externalization would in fact imply the presence of a point of symmetry at syntactic level which, in turn, would go against the assumption that c-command is a total relation.

In conclusion, Donati & Branchini (2013) offer an intriguing perspective on the relationship intercurring between language faculty and linearization requirements. By providing an unprecedented testing ground for former and recent syntactic theories dealing with word order variation, this experimental work provides a striking argument for the idea that, in the spirit of Berwick & Chomsky (2011, 2016), surface linear order is autonomous from narrow syntax rather than being directly dependent on some underlying order, being this latter either base-generated or transformationally derived from a universal phrase structure.

#### 5.4 – Conclusion

After its long history, the head-complement parameter is still a matter of discussion within the Chomskyan program. From the first proposals which were made by Graffi (1980), Stowell (1981), and Travis (1984) to most recent ones by Richards (2004, 2008) and Biberauer & Roberts (2015), it seems that the availability of such tool for accounting for the cross-linguistic variation in word order types is still being considered, if not indispensable, at least highly desirable for a linguistic theory which, like Generative Grammar, explicitly aims for explanatory adequacy.

That having been said, and coming back to the historical review proposed in this chapter, it would not be true to say that the theoretical status of this parameter has remained unscathed through an interval of more than thirty years. On the one hand, the fact that this parameter had been initially conceived as emerging from a certain degree of variation allowed by X-bar theory meant that, from a minimalist perspective, the idea of head-directionality as a parameter specified on UG was effectively doomed from the start – actually, not different from all other macroparameters which have not been reformulated in minimalistic terms. On the other hand, Kayne's (1994) Linear Correspondence Axiom



also contributed to the fall of the macroparameter of head-directionality. The idea of a linearization algorithm for syntactic terminals from which both linear order and phrase structure could be directly derived was very appealing not only for methodological reasons, as it allowed syntactic theory to dispose of a stipulative notion as head directionality at no additional cost, but also from the perspective of the Strong Minimalist Thesis, as the LCA supported the idea that the architecture of syntax is primarily driven by the need to generate linguistic structures which are optimally suited to meet interface conditions such as linearization *in primis*.

With the weakening of the LCA, which has been a consequence of the abandonment of X-bar theory in favor of Chomsky's (1995a,b) early BPS, the head-complement parameter has regained much of its former popularity within Generative Grammar. There are two leading hypotheses in this respect, which in turn can be associated with the views proposed by Biberauer & Roberts (2015) and Richards (2004, 2008) respectively. According to the account put forth by Biberauer & Roberts (2015), linear order is determined within narrow syntax, which is regarded as inherently asymmetrical (although not as rigidly asymmetrical as assumed in Kayne (1994)), with the traditional head-complement parameter being replaced by an inferential hierarchy of aggregated micro-parameters whose possible settings yield directionality effects at phrase structure level. According to Richards's (2004, 2008) proposal, on the other hand, the task of linearizing syntactic representations to comply with bare output conditions on language externalization is attributed to a revised version of Kayne's LCA which, analogously to its predecessor, applies in the PF component. The main reason for this is that, given the fact that the structure-building operation Merge is essentially symmetrical (Chomsky 2000, 2001a), syntax would nonetheless be incapable of specifying any relation of precedence between any two mutually c-commanding terminals, hence the need for a post-syntactic linearization algorithm which, differently from the LCA, is subject to parameterization.

One of the main, if not the main, advantage of Biberauer & Roberts's (2015) approach is that, in the spirit of Roberts & Holmberg's (2010) conception of parametric variation, crosslinguistic word order variation in all its possible (and impossible) combinations emerges from the interaction between the learner, the primary linguistic data, and UG, as in this model parameter setting is guided by third-factor considerations which impose to the learner to generalize a given feature of a functional head to other functional heads as long as the primary linguistic data is compatible with such generalizations. Some empirical evidence for such a hierarchical formulation of the head-

complement parameter seems to come from the FOFC, a universal constraint on phrase structure configurations which, according to the authors, can be derived straightforwardly from this very theory of linear order.

Although lacking the same range of empirical evidence that accompanies Biberauer & Roberts's (2015) proposal, Richards's (2004, 2008) formulation of the head-complement parameter adheres to the Minimalist idea that «there is a significant asymmetry between the two interfaces, with the semantic-pragmatic interface – the link to systems of thought and action – having primacy» (Berwick & Chomsky 2011, p. 30). As noted by Berwick & Chomsky (2016):

There is substantial further evidence for this conclusion. Notice again that the optimal computational operation, Merge, imposes no order on the merged elements. It follows, then, that the mental operations involving language should be independent of order, which is a reflex of the sensorimotor system. We have to impose linear order on words when we speak: the sensorimotor system does not permit production in parallel, or production of structures. The sensorimotor system was substantially in place long before language emerged, and appears to have little to do with language. (Berwick & Chomsky 2016, p. 102)

This asymmetric view of language is supported by the data collected and examined by Donati & Branchini (2013). Crucially, Donati & Branchini (2013) address the question exactly at its core by showing how blended utterances produced by bimodal children, which by definition involve the parallel activation of the acoustic and visual channels and hence of two autonomous sensori-motor spaces, can shed light on the relationship between narrow syntax and word order. In these terms, an account of word order variation along the lines of Richards (2004, 2008) has to be preferred to the one put forth in Biberauer & Roberts (2015), as the latter's strictly narrow-syntactic head-complement parameter cannot possibly comply with the fact that, in mixed and blended utterance such as those discussed by Donati & Branchini, each channel imposes its own specific linearization requirements.

## Chapter VI

### Some concluding remarks on parametrization

Since the establishment of the P&P model in Chomsky (1981a) to the advent of Minimalism, the concept of parameter has represented an optimal solution to account for cross-linguistic variation without raising a tension between descriptive and explanatory adequacy. At least initially, the idea that all possible linguistic variation could be directly accounted for by an overspecified UG specifying in advance all possible options selectable by any grammatical system had an undeniable appeal not only for its descriptive power, but also for its theory-internal contribution to the study of language faculty. After Rizzi (1978) and Taraldsen's (1978) results, comparative studies could in fact aim for the first time at shedding light on both cross-linguistic differences and, parallelly, the modules of UG which were held to be the locus of now abandoned macroparameters.

However, this picture changed well before Minimalism. After the proposal of the first microparameters, Generative Grammar naturally shifted the locus of parametric variation from universal principles to the (functional) lexicon. For the first time, therefore, systematic linguistic variation was provided with an actual format in the form of binary features. Moreover, the fact that such features were assumed to be specified on functional heads allowed the language acquisition process to be systematized without any additional postulation: parameters were in fact reduced to (a part of) the lexicon, that is, «that component of grammar for which there is strong evidence of learning» (Borer 1984 [1983], p. 29).

With the advent of the Minimalism Program, the picture changed again. Although macroparameters were clearly doomed from the start as they relied on the idea of an overspecified UG, even the first microparameters did not find a place within Minimalism. In addition to being empirically inadequate, they often seemed to be arbitrary and too much specific. In some of their earlier instantiations, they were neither binary nor small in number and, worst of all, they could be seamlessly substituted with language-particular rules.

After a period of crisis, which peaked with Newmeyer's criticism of the P&P model (Newmeyer 2004, 2005), Generative Grammar has now managed to reconceptualize the notion of parameter in a way that is compatible with the Minimalist Program. This is in large part due to Roberts & Holmberg's (2010) hierarchical parametric model, which has the merit of deriving parameters themselves, their acquisitional patterns, and their

clusterized effects from the interaction of a minimally specified UG, third factor inferential patterns and primary linguistic data. However, most interestingly for the purpose of the present thesis is that, at this theoretical stage, another typology of parametric variation has emerged from the interaction of the three factors of language design of Chomsky (2005). Assuming a completely minimal, and thus invariant, UG, the role of determining linguistic variation can in principle be shifted from the second factor, that is, unvalued features specified on functional heads, to third factor considerations, which in turn can be thought of as having a disambiguating effect on a specific set of syntactic representations which represent points of indeterminacy and, as such, cannot meet the specific bare output conditions.

As far as this latter kind of parametric variation is concerned, some ex-macroparameters like the overt vs covert movement parameter and the head-complement parameter fit nicely into this characterization. This is especially evident for cross-linguistic variation in word order patterns since, as argued by Richards (2004, 2008), syntactic structures are inherently binary and hence unfit for being decoded and externalized by the sensori-motor system. Although this same reasoning does not apply to observed duality of overt vs. covert movement, as symmetry is not involved in the specific case of wh-dislocation, if Richards N. (2010) or an equivalent PF-based account is on the right track this would nonetheless suggest «a further possible source for variability in a minimalist system: an underspecified UG relying on third-factor principles yield[ing] points of indeterminacy where the system “no longer cares”, owing to a loss of information or a lack of specification» (Richards 2008, p. 153). Crucially, there is a nontrivial difference between “PF-parameters” like head-directionality and wh-movement and “syntactic parameters” like the null subject parameter, the V-to-T movement parameter and the polysynthesis parameter: while the latter involve head-movement, the former involve movement of maximal projections. In this sense, it seems that only in narrow syntax heads can optionally move, with XPs being linearized post-syntactically. Obviously, as the remaining parameters which are now being compared are only five, this hypothesis needs to be tested against a more broad and varied body of evidence. However, this does not change the fact that this aspect contributes to further setting apart these two kinds of parametric variation.

Finally, there is one more point that deserves mention before concluding this historical review of the notion of parameter. Regardless of the conclusions that one might draw from the development of the parametric model of linguistic variation which has been

outlined in this thesis, one point which seems to be undeniable is that the change in the notion of parameter is a direct reflection of the change of the programmatic aim of the Chomskyan program. More precisely, if on the one hand the birth of the concept of parameter can be seen as the answer to Generative Grammar's need to reconcile linguistic description with explanatory adequacy, which is in turn achieved «when a descriptively adequate analysis is completed by a plausible hypothesis on its acquisition», on the other hand the latest developments of this notion, including in particular the formulation of PF-based parameters, would have not been reached if generative theory's new programmatic aim had not been to investigate how well designed language faculty is relative to interface conditions (Chomsky 2002, p. 129).



## REFERENCES

- ABELS, Klaus and NEELEMAN, Ad (2007) "Linear asymmetries and the LCA". Ms, University College London.
- AOUN, Joseph and SPORTICHE, Dominique (1981) "On the formal theory of government". *The Linguistic Review* 2: 211-236.
- ARREGI, Karlos (2002) *Focus on Basque Movements*. PhD dissertation, MIT.
- BACH, Emmon, and HORN, George M. (1976). "Remarks on Conditions on Transformations". *Linguistic Inquiry* 7: 265-299.
- BAKER, Mark (1970) "Notes on the description of English questions: the role of an abstract question morpheme". *Foundations of Language* 6: 197-219.
- BAKER, Mark (1988) *Incorporation: A Theory of Grammatical Function Changing*. Chicago: Chicago University Press.
- BAKER, Mark (1996) *The Polysynthesis Parameter*. New York: Oxford University Press.
- BAKER, Mark (2001) *The Atoms of Language*. New York: Basic Books.
- BAKER, Mark (2008a) "The macroparameter in a microparametric world". Ms, Rutgers University. Subsequently published in *The Limits of Syntactic Variation* ed. by T. Biberauer. Amsterdam: John Benjamins, 2008.
- BAKER, Mark (2008b) *The Syntax of Agreement and Concord*. Cambridge: Cambridge University Press.
- BAKER, Mark (2011) *Principles and Parameters Set Out from Europe*. Handout, MIT Linguistics 50th Anniversary, 9 December 2011.

- BERWICK, Robert C. and CHOMSKY, Noam (2011) "The biolinguistic program: the current state of its development". *The Biolinguistic Enterprise* ed. by A. Di Sciullo & C. Boeckx, 19-41. Oxford: Oxford University Press.
- BERWICK, Robert C. and CHOMSKY, Noam (2016) *Why Only Us*. Cambridge, MA: MIT Press.
- BIBERAUER, Theresa (2008) "Introduction". *The Limits of Syntactic Variation* ed by T. Biberauer, 1-72. Amsterdam: Benjamins.
- BIBERAUER, Theresa, HOLMBERG, Anders and ROBERTS, Ian (2009) "Linearization and the architecture of grammar: a view from the Final-over-Final Constraint". *Studies in Linguistics, Volume 3 (Proceedings of XXXV Incontro di Grammatica Generativa)* ed. by V. Moscati & E. Servidio, 77-89. Siena: CISCL Press.
- BIBERAUER, Theresa, HOLMBERG, Anders and ROBERTS, Ian (2014) "A syntactic universal and its consequences". *Linguistic Inquiry* 45: 169-225.
- BIBERAUER, Theresa, HOLMBERG, Anders, ROBERTS, Ian and SHEEHAN, Michelle (2014) "Complexity in comparative syntax: the view from modern parametric theory". *Measuring Grammatical Complexity* ed. by F. Newmeyer & L. Preston, 103-127. Oxford: Oxford University Press.
- BIBERAUER, Theresa, NEWTON, Glenda and SHEEHAN, Michelle (2009) "Limiting synchronic and diachronic variation and change: the Final-Over-Final Constraint". *Language and Linguistics* 10: 699-741.
- BIBERAUER, Theresa, NEWTON, Glenda and SHEEHAN, Michelle (2010) "Impossible changes and impossible borrowings". *Continuity and Change in Grammar* ed. by A. Breitbarth, C. Lucas, S. Watts & D. Willis, 35-60. Amsterdam: Benjamins.
- BIBERAUER, Theresa and ROBERTS, Ian (2010) "Subjects, tense and verb-movement". *Parametric Variation: null subjects in Minimalist Theory* ed. by T. Biberauer, A.



Holmberg, I. Roberts & M. Sheehan, 263-302. New York: Cambridge University Press.

BIBERAUER, Theresa and ROBERTS, Ian (2012) "Towards a parameter hierarchy for auxiliaries: diachronic considerations". *Cambridge Occasional Papers in Linguistics* 6: 267-294.

BIBERAUER, Theresa and ROBERTS, Ian (2015) "The clausal hierarchy, features, and parameters". *Beyond Functional Sequence: The Cartography of Syntactic Structures, Volume 10* ed. by U. Shlonsky, 296-313. New York: Oxford University Press.

BOECKX, Cedric (2008) "Approaching parameters from below". *The Biolinguistics Enterprise: New perspectives on the evolution of the human language faculty* ed. by A. M. Di Sciullo & C. Boeckx, 205-221. Oxford: Oxford University Press.

BOECKX, Cedric (2010) "What principles and parameters got wrong". Ms, ICREA/UAB. Available online at [<http://ling.auf.net/lingbuzz/001118>]. Subsequently published in *Linguistic Variation in the Minimalist Framework* ed. by C. Picallo, 13-35. Oxford: Oxford University Press, 2014.

BOECKX, Cedric and KLEANTHES, K. Grohmann (2004) "Barriers and phases: forward to the past?". Paper presented at TiLT 2004, Budapest, May 2004.

BORER, Hagit (1984) *Parametric Syntax*. Dordrecht: Foris Publications. First edition 1983.

BRESNAN, Joan W. (1970) "On complementizers: toward a syntactic theory of complement types". *Foundations of Language* 6: 297-321. Cambridge, MA: MIT Press.

BRESNAN, Joan W. (1982) "Control and complementation". *Linguistic Inquiry* 13: 343-434.

BURZIO, Luigi (1986) *Italian Syntax: a Government Binding Approach*. Dordrecht: Reidel.

- CHOMSKY, Noam (1957) *Syntactic Structures*. The Hague: Mouton.
- CHOMSKY, Noam (1965) *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- CHOMSKY, Noam (1970) "Remarks on nominalization". *Readings in English Transformational Grammar* ed. by R. A. Jacobs & P. S. Rosenbaum, 184-221. Boston: Ginn.
- CHOMSKY, Noam (1973) "Conditions on Transformations". *A Festschrift for Morris Halle* ed. by S. Anderson & P. Kiparsky, 232-286. New York: Holt, Rinehart and Winston. (Reprinted in Chomsky 1977b. 79-160).
- CHOMSKY, Noam (1975a) [1955-56] *The Logical Structure of Linguistic Theory*. Chicago & London: The University of Chicago Press.
- CHOMSKY, Noam (1975b) *Reflections on Language*. New York: Random House.
- CHOMSKY, Noam (1976) "Conditions on Rules of Grammar". *Linguistic Analysis* 2: 303-351. (Reprinted in Chomsky 1977b, 163-210).
- CHOMSKY, Noam (1977a) "On wh-movement". *Formal Syntax* ed. by P. Culicover, A. Akmajian & T. Wasow, 71-132. New York: Academic Press.
- CHOMSKY, Noam (1977b) *Essays on Form and Interpretation*. New York-Amsterdam-Oxford: North Holland.
- CHOMSKY, Noam (1980) "On binding". *Linguistic Inquiry* 11: 1-46.
- CHOMSKY, Noam (1981a) *Lectures on Government and Binding: the Pisa Lectures*. Holland: Foris Publications. Reprint, Berlin: Mouton de Gruyter, 1993.
- CHOMSKY, Noam (1981b) "Principles and parameters in syntactic theory". *Explanation in Linguistics* ed. by N. Hornstein & D. Lightfoot, 32-75. London & New York: Longman.

- CHOMSKY, Noam (1982) *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA: MIT Press.
- CHOMSKY, Noam (1986a). *Barriers*. Cambridge, MA: MIT Press.
- CHOMSKY, Noam (1986b) *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.
- CHOMSKY, Noam (1992) "A minimalist program for linguistic theory". *MIT Occasional Papers in Linguistics* 1. Reprinted as Chapter 3 of Chomsky (1995a), 167-217.
- CHOMSKY, Noam (1995a) *The Minimalist Program*. Cambridge, MA: MIT Press.
- CHOMSKY, Noam (1995b) "Bare phrase structure". *Evolution and Revolution in Linguistic Theory: A Festschrift in Honor of Carlos Otero* ed. by H. Campos & P. Kempchinsky, 51-109. Washington, DC: Georgetown University Press.
- CHOMSKY, Noam (2000) "Minimalist inquiries: the framework". *Step By Step. Essays on Minimalist Syntax in Honor of Howard Lasnik* ed. by R. Martin, D. Michaels & J. Uriagereka, 89-155. Cambridge, MA: MIT Press.
- CHOMSKY, Noam (2001a) "Beyond explanatory adequacy". *MIT Occasional Papers in Linguistics* 20: 1-28.
- CHOMSKY, Noam (2001b) "Derivation by phase". *Ken Hale: A Life in Language* ed. by M. Kenstowicz, 1-52. Cambridge, MA: MIT Press.
- CHOMSKY, Noam (2002) "An interview on Minimalism". *On Nature and Language*, ed. by A. Belletti & L. Rizzi, 92-161. Cambridge: Cambridge University Press.
- CHOMSKY, Noam (2004) "Beyond explanatory adequacy". *Structures and Beyond: The Cartography of Syntactic Structures, Volume 3* ed. by A. Belletti, 104-131. Oxford, Oxford University Press.

- CHOMSKY, Noam (2005) "Three factors in language design". *Linguistic Inquiry* 36: 1-22.
- CHOMSKY, Noam (2008) "On phases". *Foundational Issues in Linguistic Theory* ed. by R. Freidin, C. P. Otero & M. L. Zubizarreta, 133-166. Cambridge, MA: M.I.T. Press.
- CHOMSKY, Noam and LASNIK, Howard (1977) "Filters and control". *Linguistic Inquiry* 8: 425-504.
- DEN BESTEN, Hans (1981) "On the interaction of root transformations and lexical deletive rules". *Groninger Arbeiten zur germanistischen Linguistik* 20.
- DEN BESTEN, Hans (1989) "On the interaction of root transformations and lexical deletive rules". *Studies in West Germanic Syntax, Dissertatie Tilburg* 20: 14-100. Formerly published in *On the Formal Syntax of Westgermania* ed. by W. Abraham, Amsterdam: Benjamins, 1983.
- DIESING, Molly (1990) "Verb movement and the subject position in Yiddish". *Natural Language and Linguistic Theory* 8: 41-80.
- DONATI, Caterina and BRANCHINI, Chiara (2013). "Challenging linearization: simultaneous mixing in the production of bimodal bilinguals". *Challenges to Linearization* ed. by T. Biberauer & I. Roberts, 93-128. Berlin, Germany: Mouton de Gruyter.
- DOWNING, Laura (2005) "The prosody of focus-related enclitics in some Southern Bantu languages". Handout from talk given at SOAS.
- DRYER, Matthew (2009) "The branching direction theory revisited". *Universals of Language Today* ed. by S. Scalise, E. Magni & A. Bisetto, 185-207. Berlin: Springer.
- EMMOREY, Karen, BORINSTEIN, Helsa and THOMPSON, Robin (2005) "Bimodal bilingualism: code blending between spoken English and ASL". *Proceedings of International Symposium on Bilingualism 4* ed by J. Cohen, K. McAlister, K. Rolstad & J. MacSwan, 663-673. Somerville, MA: Cascadilla.

- EMONDS, Joseph (1976) *A Transformational Approach to English Syntax: Root, Structure-Preserving, and Local Transformations*. New York: Academic Press.
- EMONDS, Joseph (1978) "The verbal complex V'-V in French". *Linguistic Inquiry* 9: 151-175.
- EPSTEIN, Samuel David, GROAT, Erich M., KAWASHIMA, Ruriko and HISATSUGU, Kitahara (1998) *A Derivational Approach to Syntactic Relations*. New York: Oxford University Press.
- FALK, Cecilia (1993) *Non-referential Subjects in the History of Swedish*. PhD dissertation, Department of Scandinavian languages, University of Lund.
- FRASCARELLI, Mara (2007) "Subjects, topics, and the interpretation of referential pro". *Natural Language and Linguistic Theory* 25: 691-734.
- GEORGE, Leland M. (1980) *Analogical Generalization in Natural Language Syntax*. Ph.D. dissertation, MIT.
- GIANOLLO, Chiara, GUARDIANO, Cristina and LONGOBARDI, Giuseppe (2008) "Three fundamental issues in parametric linguistics". *The Limits of Syntactic Variation* ed. by T. Biberauer, 109-42. Amsterdam: Benjamins.
- GILLIGAN, Gary M. (1987) *A Cross-Linguistic Approach to the Pro-Drop Parameter*. Ph.D. dissertation, University of Southern California.
- GIVÓN, Talmy (1983) "Topic continuity in discourse: an introduction". *Topic Continuity in Discourse: A Quantitative Cross-language Study* ed. by T. Givón, 5-41. Amsterdam/Philadelphia: John Benjamins.
- GRAFFI, Giorgio (1980) "'Universali di Greenberg' e grammatica generativa". *Lingua e Stile* 15: 371-387.

- GRAFFI, Giorgio (2001) *200 Years of Syntax: A Critical Survey*. Amsterdam-Philadelphia: John Benjamins Publishing Company.
- GRAFFI, Giorgio (2008) *Che cos'è la Grammatica Generativa*. Roma: Carocci.
- GRAFFI, Giorgio (2010) *Due Secoli di Pensiero Linguistico*. Roma: Carocci.
- GRAFFI, Giorgio and RIZZI, Luigi, eds. (1979) *La Sintassi Generativo-Trasformatzionale*. Bologna: Il Mulino.
- GREENBERG, Joseph H. (1963) "Some universals of grammar with particular reference to the order of meaningful elements". *Universals of Language* ed. by Joseph H. Greenberg, 58-90. London: MIT Press.
- GRIMSHAW, Jane (1981) "Subcategorization of grammatical relations". Paper presented at the Harvard Mini-Conference on the Representation of Grammatical Relations, Harvard University, December 12, 1981.
- GRIMSHAW, Jane (2000) "Locality and extended projection". *Lexical Specification and Insertion* ed. by P. Coopmans, M. Everaert & J. Grimshaw, 115-133. Amsterdam: John Benjamins.
- GUMPERZ, John (1982) *Discourse strategies*. Cambridge: Cambridge University Press.
- HALE, Kenneth (1980) "Remarks on Japanese phrase structure". *MIT Working Papers in Linguistics* 2: 185-203.
- HALE, Kenneth (1981) [1979] "On the Position of Warlpiri in a Typology of the Base". Bloomington: Indiana University Linguistics Club.
- HALE, Kenneth (1983) "Warlpiri and the grammar of non-configurational languages". *Natural Language and Linguistic Theory* 1: 5-47.

- HALE, Kenneth (1986) "On nonconfigurational structures". *Anuario del Seminario de Filología Vasca "Julio de Urquijo"* ASJU XX-2, 351-358.
- HALE, Kenneth (1989) "On nonconfigurational structures". *Configurationality: the Typology of Asymmetries* ed. by L.K. Marácz & P. Muysken, 293-300. Dordrecht: Foris.
- HASPELMATH, Martin, DRYER, Matthew S., GIL, David and COMRIE, Bernard, eds. (2008) *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library.
- HAUSER, Marc, CHOMSKY, Noam and FITCH, Tecumseh (2002) "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?". *Science* 298: 1569-1579.
- HOLMBERG, Anders (2005) "Is there a little *pro*? Evidence from Finnish". *Linguistic Inquiry* 36: 533-564.
- HOLMBERG, Anders (2010a) "Null subject parameters". *Parametric Variation: Null Subjects in Minimalist Theory* ed. by T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan, 88-124. New York: Cambridge University Press.
- HOLMBERG, Anders (2010b) "Parameters in minimalist theory: the case of Scandinavian". *Theoretical Linguistics* 36.1: 1-48.
- HOLMBERG, Anders and ROBERTS, Ian (2013) "The syntax–morphology relation". *Lingua* 130, 111-131.
- HORNSTEIN, Norbert and WEINBERG, Amy (1981) "Case theory and preposition stranding". *Linguistic Inquiry* 12: 55-91.
- HORNSTEIN, Norbert, JAIRO, Nunes and KLEANTHES, Grohmann (2005) *Understanding Minimalism*. Cambridge: Cambridge University Press.

- HUANG, James. (1982) *Logical Relations in Chinese and the Theory of Grammar*. PhD dissertation, MIT.
- HUANG, James, AUDREY, Li and YAFEI, Li (2009) *The Syntax of Chinese*. Cambridge University Press.
- ISHIHARA, Shinichiro (2003) *Intonation and Interface Conditions*. PhD dissertation, MIT.
- JACKENDOFF, Ray (1977) *X' Syntax: a Study of Phrase Structure*. Cambridge MA: MIT Press.
- JOSEPH, Brian (1976) "Raising in Modern Greek: A Copying Process?" In *Harvard Studies in Syntax and Semantics* 2: 241-278.
- KAYNE, Richard (1981) "Two notes on the NIC". *Theory of Markedness in Generative Grammar: Proceedings of the 1979 GLOW Conference*, ed. by A. Belletti, L. Brandi & L. Rizzi, 317-346. Pisa: Scuola Normale Superiore di Pisa. Reprinted as Chapter 2 of Kayne (1983), 23-46.
- KAYNE, Richard (1983) *Connectedness and Binary Branching*, Dordrecht: Foris Publications.
- KAYNE, Richard (1994) *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
- KAYNE, Richard (2000) "Microparametric syntax: some introductory remarks". *Parameters and Universals* ed. by R. Kayne, 3-9. New York: Oxford University Press.
- KAYNE, Richard (2005) "Some notes on comparative syntax, with special reference to English and French". *The Oxford Handbook of Comparative Syntax* ed. by G. Cinque & R. Kayne, 3-69. New York: Oxford University Press.
- KISS, Katalin (1995) "Introduction". *Discourse configurational languages*, ed. by K. É. Kiss, 3-27. Oxford: Oxford University Press.



- KOENEMAN, Olaf and NEELEMAN, Ad (2001) "Predication, verb movement and the distribution of expletives". *Lingua* 111: 189-233.
- KOOPMAN, Hilda and SPORTICHE, Dominique (1991) "The Position of Subjects". *Lingua*, 85: 211-258.
- LI, Charles N. and THOMPSON, Sandra A. (1973) "Historical change of word order: a case study in Chinese and its implications". *Historical Linguistics: Proceedings of the First International Conference on Historical Linguistics, Edinburgh, 2nd-7th September 1973* ed. by J. M. Anderson & C. Jones, 199-217. North-Holland.
- LI, Charles N. and THOMPSON, Sandra A. (1975) "The semantic function of word order: A case study in Mandarin". *Word Order and Word Order Change* ed. by C. Li, 163-197. Austin: University of Texas Press.
- LIGHT, Timothy (1979) "Word order and word order change in Mandarin Chinese". *Journal of Chinese Linguistics* 7: 149-180.
- MANZINI, Maria Rita and WEXLER, Kenneth (1987) "Parameters, binding, and learning theory". *Linguistic Inquiry* 18, 413-444.
- MASSAM, Diane (2000) "VSO and VOS: aspects of Niuean word order". *The Syntax of Verb-Initial Languages*, ed. by A. Carnie & E. Guilfoyle, 97-116. Oxford University Press, New York.
- MASSAM, Diane (2005). "Lexical categories, lack of inflection, and predicate fronting in Niuean". *Verb First: On the Syntax of Verb-Initial Language* ed. by A. Carnie, S. Dooley & H. Harley, 227-242. Amsterdam: Benjamins.
- MAY, Robert (1977) *The Grammar of Quantification*. PhD dissertation, MIT.
- MAY, Robert (1985) *The Logical form*. Cambridge, MA: MIT Press.

- McCLOSKEY, James (1996) "On the scope of verb-movement in Irish". *Natural Language and Linguistic Theory* 14: 47-104.
- MORO, Andrea (1988) "Per una teoria unificata delle frasi copulari". *Rivista di Grammatica Generativa* 13: 81-110.
- NEWMAYER, Frederick J. (2004) "Against a parameter-setting approach to language variation". *Language Variation Yearbook* 4: 181-234.
- NEWMAYER, Frederick J. (2005) *Possible and Probable Languages*. Oxford: Oxford University Press.
- NEWTON, Glenda (2008) "Final C elements: complementisers and particles". Handout, University of Cambridge.
- NIKITINA, Tatiana (2008) "Nominalisation and word-order change in Niger-Congo". Unpublished Ms: Stanford University.
- NUNES, Jairo (2004) *Linearization of Chains and Sideward Movement*. Cambridge, MA: MIT Press.
- POLLARD, Carl and SAG, Ivan (1992) "Anaphors in English and the scope of binding theory". *Linguistic Inquiry* 23: 261-303.
- POLLOCK, Jean-Yves (1989) "Verb movement, universal grammar, and the structure of IP". *Linguistic Inquiry* 20: 365-424.
- POSTAL, Paul (1962) *Some Syntactic Rules in Mohawk*. PhD dissertation, Yale University, New Haven, Connecticut. Subsequently published by Garland Press, New York, 1979.
- REINHART, Tanya (1976) *The Syntactic domain of Anaphora*. PhD dissertation, MIT.

- REINHART, Tanya (1981) "Pragmatics and linguistics: an analysis of sentence topics". *Philosophica* 27: 53-94.
- REINHART, Tanya and REULAND, Eric (1993) "Reflexivity". *Linguistic Inquiry* 24: 657-720.
- RICHARDS, Marc (2004) *Object Shift and Scrambling in North and West Germanic: A Case Study in Symmetrical Syntax*. PhD dissertation, University of Cambridge.
- RICHARDS, Marc (2008) "Two kinds of variation in a Minimalist system". *Linguistische Arbeits Berichte* 87: 133-162.
- RICHARDS, Norvin (2010) *Uttering Trees*. Cambridge, MA: MIT Press.
- RIZZI, Luigi (1978) "Violations of the wh island constraint in Italian and the subjacency condition". *Montreal Working Papers in Linguistics* 11. Reprinted as Chapter 2 of Rizzi (1982), 49-116.
- RIZZI, Luigi (1982) *Issues in Italian Syntax*. Dordrecht: Foris Publications.
- RIZZI, Luigi (1997) "The fine structure of the left periphery". *Elements of Grammar* ed. by L. Haegeman, 281-337. Dordrecht: Kluwer.
- RIZZI, Luigi (2001) "Relativized Minimality effects". *The Handbook of Contemporary Syntactic Theory* ed. by M. Baltin & C. Collins, 89-110. Oxford: Blackwell.
- RIZZI, Luigi (2013) "Theoretical and comparative syntax: some current issues". *L'Interface Langage-Cognition: actes du 19e Congrès International des Linguistes: Genève, 22-27 juillet 2013* ed. by S. Anderson, J. Moeschler & F. Reboul, 307-331. Genève: Droz.
- RIZZI, Luigi (2014) "On the elements of syntactic variation". *Linguistic Variation in the Minimalist Framework* ed. by C. Picallo, 13-35. Oxford: Oxford University Press.

- ROBERTS, Ian (2001) "Language change and learnability". *Language Acquisition and Learnability* ed. by S. Bertolo, 81-125. Cambridge: Cambridge University Press.
- ROBERTS, Ian (2004) "The C-system in Brythonic Celtic languages". *The Structure of CP and IP: The Cartography of Syntactic Structures, Volume 2* ed. by L. Rizzi, 297-328. New York and Oxford: Oxford University Press.
- ROBERTS, Ian (2007) *Diachronic Syntax*. Oxford: Oxford University Press.
- ROBERTS, Ian (2010) "A deletion analysis of null subjects". *Parametric Variation: null subjects in Minimalist Theory* ed. by T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan, 58-87. New York: Cambridge University Press.
- ROBERTS, Ian (2012) "Macroparameters and minimalism: a programme for comparative research". *Parametric Theory and Linguistic Change* ed. by C. Galves, S. Cyrino, R. Lopez & J. Avelar, 320-335. Oxford: Oxford University Press.
- ROBERTS, Ian (2016) "Some remarks on parameter hierarchies". *Rethinking Parameters* ed. by L. Eguren, O. Fernandez-Soriano & A. Mendikoetxea, 170-200. Oxford: Oxford University Press.
- ROBERTS, Ian and HOLMBERG, Anders (2005) "On the role of parameters in Universal Grammar: a reply to Newmeyer". *Organizing Grammar: Linguistic Studies in Honor of Henk van Riemsdijk* ed. by H. Broekhuis, N. Corver, R. Huybregts, U. Kleinhenz & J. Koster, 538-553. Berlin: Mouton de Gruyter.
- ROBERTS, Ian and HOLMBERG, Anders (2010) "Introduction: parameters in minimalist theory". *Parametric Variation: Null Subjects in Minimalist Theory* ed. by T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan, 1-57. New York: Cambridge University Press.
- ROSS, John Robert (1986) *Infinite Syntax!*. Norwood, N.J.: Ablex. Formerly published as *Constraints on Variables in Syntax*. PhD dissertation, MIT, 1967.

- SAITO, Mamoru (1982) "Scrambling, topicalization, and strong crossover". Ms, MIT, Cambridge, MA.
- SAFIR, Kenneth J. (1985) *Syntactic Chains*. Cambridge: Cambridge University Press.
- SAMEK-LODOVICI, Vieri (1996) *Constraints on Subjects, an Optimality Theoretic Analysis*. PhD dissertation, Rutgers University.
- SPORTICHE, Dominique (1981) "Bounding nodes in French" *The Linguistic Review* 1: 219-246.
- SPORTICHE, Dominique (1988) "Conditions on silent categories". Ms, UCLA.
- SPORTICHE, Dominique (1990) "Movement, agreement, constituent structure and case" (provisional title). Ms, UCLA.
- SPORTICHE, Dominique (2013) "Binding theory–structure sensitivity of referential dependencies". *Lingua* 130: 187-208.
- STOWELL, Timothy A. (1981) *Origins of Phrase Structure*. PhD dissertation, MIT, Cambridge, MA.
- TALLERMAN, Maggie (1996) "Fronting constructions in Welsh". *The Syntax of the Celtic Languages* ed. by R. Borsley & I. Roberts, 97-124. Cambridge: Cambridge University Press.
- TARALDSEN, Knut Tarald (1978) "On the NIC, vacuous application and the that-trace filter". Bloomington: Indiana University Linguistics Club.
- TRAVIS, Lisa (1984) *Parameters and Effects of Word Order Variation*. PhD dissertation, MIT, Cambridge, MA.
- TRITHART, Mary (1977) *Relational Grammar and Chichewa Subjectivization*. PhD dissertation, UCLA, Los Angeles.

- VAN RIEMSDIJK, Henk (1978) *A Case Study in Syntactic Markedness*. Lisse: Peter de Ridder Press.
- WHITMAN, John (1982) "Configurationality parameters". Ms, Department of Linguistics, Harvard University.
- WILLIAMS, Edwin (1981) "Argument structure and morphology". *The Linguistic Review* 1: 81-114.
- WOODBURY, Hanni (1975) *Noun Incorporation in Onondaga*. PhD dissertation, Yale University, New Haven, Connecticut.